

72798

mg

184368

STIC-Biotech/ChemLib

From: Portner, Ginny
Sent: Wednesday, April 05, 2006 2:30 PM
To: STIC-Biotech/ChemLib
Subject: 10/768,093

please search SEQ Id No 9 and 4 with respect to polypeptide/amino acid sequences.

Ginny Portner
Remsen Building
Art Unit 1645
Room E03, B02; Mail Box 3C18
(571) 272-0862

RECEIVED
APR - 5 2006
.../CHEM LIB
(STIC)


```

F6A1_ECOLI STANDARD: PRT: 154 AA.
ID F6A1_ECOLI
AC P53508.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE CS6 fimbrial subunit A precursor (CS6 pilin).
DB CS6 fimbrial subunit A precursor (CS6 pilin).
GN Name=csaA;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
XN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=O167:H5 / E10703 / EIEC;
RA Wolf M.K., de Haan L., Cassels F.C., Willshaw G.A., van Gestel E.,
RA Gaastera W., Warren R., Boedeker E.C.;
RL Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Fimbriae (also called pili), polar filaments radiating
CC from the surface of the bacterium to a length of 0.5-1.5
CC micrometers and numbering 100-300 per cell, enable bacteria to
CC colonize the epithelium of specific host organs.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL, U04844; AAC45093.1; -; Unassigned_DNA.
DR PIR; I60286; I60266.
KW Fimbria, signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 154 CS6 fimbrial subunit A.
SQ SEQUENCE 154 AA; 16953 MW; B4801BBE6B6BD206 CRC64;
-----
5 try Match 93.7%; Score 655; DB 1; Length 154;
Blast Local Similarity 91.9%; Pred. No. 1.4e-52;
Matches 125; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
QY 1 RTEIATKNFPSTTISKSPFAPEPRIQPSFGENVGKEGALFSVNLTVENVNQVTVYPV 60
DB 19 RTEIATKNFPSTTISKSPFAPEPRIQPSFGENVGKEGALFSVNLTVENVNQVTVYPV 78
QY 61 YDEBYGGLRLVNTADASQSIITYQIYDEYDGKKMLKDHGAEVTPNQGITKALNTSGEKI 120
DB 79 YDEBYGGLRLVNTADDSQSIITYQIYVDGGRKMLKDHGAEVTPNQGITRALNTSGEKI 138
QY 121 SPGIYDQVMVGYGVN 136
DB 139 PPGIYDQVMVGYGVN 154
-----
RESULT 3
F6B1_ECOLI STANDARD: PRT: 167 AA.
ID F6B1_ECOLI
AC P53510.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE CS6 fimbrial subunit B precursor.
DB CS6 fimbrial subunit B precursor.
GN Name=csaB;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
XN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=O167:H5 / E10703 / EIEC;
RA Wolf M.K., de Haan L., Cassels F.C., Willshaw G.A., van Gestel E.,
RA Gaastera W., Warren R., Boedeker E.C.;
RL Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC -----

```

```

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U04844; AAC45094.1; -; Unassigned_DNA.
KW PIR; I83348; I83348.
KW Fimbrria; Signal.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 167 CS6 fimbrial subunit B.
SQ SEQUENCE 167 AA; 18022 MW; F9FPIB9IC411848E CRC64;
Query Match 17.2%; Score 120.5; DB 1; Length 167;
Best Local Similarity 30.1%; Pred. No. 0.0047;
Matches 43; Conservative 24; Mismatches 61; Indels 15; Gaps 7;
QY 7 KNFPEVSTTIKSPFAPE---PRIQPSFGENVKGEKALLFSVNLTVPENVSQTVVP-- 59
DQ 27 KSLDVVNINIEQNFI-PDIDSAVAILIPNVYSDPKLSQLYVEMTIPAGSAVKIVPTDS 85
QY 60 VYDEVDYGLGLVNTADASQSIITQIYDER--GKKMLKDHA-EVTPNOQTTFALANTSG 116
DQ 86 LTSSGQOIGLTVANNPDDQNNMYIRKDSGAGRFMAGQKGSFVKENTSYTFSAI-YTGG 144
QY 117 E---KKISPGIYNDQVWVGYYVN 136
DQ 145 EYPNSGYSGTGYAGHLTVSPYSN 167
RESULT 4
F6B2_ECOLI STANDARD; PRT; 167 AA.
ID F6B2_ECOLI
AC P53511;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE CS6 fimbrial subunit B precursor.
GN Name=cs6B;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NC NCB1_Textid=562;
RX [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=B8775;
RC Wolf M.K., de Haan L.A.M., Cassels F.C., Willschaw G.A., Geesbel E.C.M.,
RA Gaastera W., Warren R., Boedeker E.C.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U04846; AAB51362.1; -; Unassigned_DNA.
KW Fimbrria; Signal.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 167 CS6 fimbrial subunit B.
SQ SEQUENCE 167 AA; 17930 MW; BBD0D041D3BBFBDCB CRC64;
Query Match 16.2%; Score 113.5; DB 1; Length 167;
Best Local Similarity 28.5%; Pred. No. 0.021;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;
QY 7 KNFPEVSTTIKSPFAPE---PRIQPSFGENVKGEKALLFSVNLTVPENVSQTVVP-- 59
DQ 27 KSLDVVNINIEQNFI-PDIDSAVAILIPNVYSDPKLSQLYVEMTIPAGSAVKIVPTDS 85
QY 60 VYDEVDYGLGLVNTADASQSIITQIYDER--GKKMLKDHA-EVTPNOQTTFALANTSG 108
DQ 86 LTSSGQOIGLTVANNPDDQNNMYIRKDSGAGRFMAGQKGSFVKENTSYTF 137

```

QY 109 KALNTSGE---KKISPGIYNDQVMGVYVN 136
DB 138 SAI-YTGGEYPNSGSGTYAGNLTVSFYN 167
RESULT 5
ID 054U88_DICDI PRELIMINARY; PRT; 440 AA.
AC 054U88;
DT 13-SEP-2005 (TEMBLrel. 31, Created)
DT 13-SEP-2005 (TEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
ORFNames=DD80215209;
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
NCBI_TaxID=44689;
RN 11
RC NUCLEOTIDE SEQUENCE.
RA Etlinger L., Pachbat J.A., Gloeckner G., Rajandream M.-A.,
RA Tunggal B., Bertman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Drieseche N., Cronin A., Goodhead I.,
RA Muzny D., Moutier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchteser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Louieged H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Urushihara H., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Ma J., Kohara Y., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstein G., Rosenblatt A., Cox E.C.,
RA Chishti R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.A.,
RA Williams J., Dear P.H., Noegel A.A., Barrett B., Kuspa A.,
RA The genome of the social amoeba Dictyostellium discoideum",
RT Nature 0:0-0(2005).
RL -1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL: AAFI0100071; EAL66967.1; -; Genomic_DNA.
CC KW Hypothetical protein.
SQ SEQUENCE 440 AA; 50234 MW; FC85D88231E1A694 CRC64;
Query Match 13.2%; Score 92; DB 2; Length 440;
Best Local Similarity 30.3%; Pred. No. 6.4;
Matches 36; Conservative 16; Mismatches 33; Indels 34; Gaps 7;
QY 44 VNLTVBENSGVYVYVYDDEYGLRVNTPADSQ-----SIYQ-----IYDEK 88
DB 259 LNTLPLEISIOETV-----YWEPKLNTITDYQFNPNPISFDNNSDVYISIPQIDVK 312
QY 89 GKIKLKDGHGAELVPN---OOITFKALNTYS-----GEKKSIPGIYNDQ---VMGVYVN 136
DB 313 GVATIGTHSGNVLENMSESKLPFKENYDKYGVGE--YMPGLNNDKEGHSVYCHTN 369
RESULT 6
ID 072P30_LEPIC PRELIMINARY; PRT; 918 AA.
AC 072P30;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Membrane carboxypeptidase.
ORFNames=LIC12646;
OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar
Copenhagen).
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.

OX NCBI_TaxID=44275;
RN 11
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=Flacruz LI-130;
RX PubMed=15028702; DOI=10.1126/JB.186.7.2164-2172.2004;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartshorn R.A.,
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,
RA Coutinho L.L., Degraive W.M., Dellagostin O.A., El-Dorri H.,
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gambertini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,
RA Kuranee E.E., Lemos E.G.M., Lemos M.V.P., Marino C.L., Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schirfer A.,
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.D.G., Ferro J.A.,
RA Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Slyke M.A.,
RT "Comparative genomics of two Leptospira interrogans serovars reveals
RT novel insights into physiology and pathogenesis.",
RL J. Bacteriol. 186:2164-2172(2004).
DR EMBL: AEO17297; AAS71206.1; -; Genomic_DNA.
DR GO: GO:0009274; Cell wall (sensu Bacteria); IEA.
DR GO: GO:0004180; F:carboxypeptidase activity; IEA.
DR GO: GO:0003824; F:catheptic activity; IEA.
DR GO: GO:0006558; F:penicillin binding; IEA.
DR GO: GO:0009252; F:peptidoglycan biosynthesis; IEA.
DR Interpro: IPR001264; Glyco_trans_51.
DR Interpro: IPR001460; Penicil_bind_1pept.
DR Pfam: PF009912; Transglyl; 1.
DR Pfam: PF00905; Transpeptidase; 1.
DR ProDom: PD001895; Glyco_trans_51; 1.
KW Carboxypeptidase; Complete proteome.
SQ SEQUENCE 918 AA; 102807 MW; 2D9E5321B0396960 CRC64;
Query Match 12.4%; Score 87; DB 2; Length 918;
Best Local Similarity 25.9%; Pred. No. 44;
Matches 37; Conservative 25; Mismatches 57; Indels 24; Gaps 6;
QY 6 TKNPVSTTISKSPFAPDEPRIPQSPFENYKGEALLFSVNLTVBENSGVYVYDDEY 65
DB 478 TCGFSYTYTVESEVQALPRTIVAKYVNDVQKNG-LVRKTRLTNNKSSSEYAVRRYQDL 536
QY 66 -GLGLVNT-----ADASQSIYQIYDEKSKKMLKDGHAEVTPNQIIFKA---LNTY 114
DB 537 SPALBLPIPDGSGGQNESLQVALVAVDPSTGEILMHGGS-----EFKADNQDLRT 589
QY 115 SGEEK-----ISPGIYNDQVMVG 132
DB 590 TAMKROTGSIKRILYSALIEFG 612
RESULT 7
ID 08F7D7_LEPIN PRELIMINARY; PRT; 918 AA.
AC 08F7D7;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1A (EC 3.4.-.-) (EC 2.4.2.-).
GN OrderedLocuNames=LAI1009;
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN 11
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01597;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Yu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.,

"Unique physiological and pathogenic features of *Leptospira* interorgans revealed by whole-genome sequencing.";
 RT Nature 422:888-893(2003).
 DR EMBL; AE011285; AAM48208.1; -; Genomic DNA.
 DR GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0008658; F:penicillin binding; IEA.
 DR GO; GO:0009252; F:peptidoglycan biosynthesis; IEA.
 DR InterPro; IPR001264; Glyco_trans_51.
 DR InterPro; IPR001460; Penic_bind_type.
 DR Pfam; PF00912; Transgly_1.
 DR Pfam; PF00905; Transpeptidase; 1.
 DR ProDom; PD001895; Glyco_trans_51; 1.
 DR Complete proteome.
 SO SEQUENCE 918 AA; 102818 MW; 5EEAFD48DC58326 CRC64;
 Query Match 12.4%; Score 87; DB 2; Length 918;
 Best Local Similarity 25.9%; Pred. No. 44;
 Matches 37; Conservative 25; Mismatches 57; Indels 24; Gaps 6;
 QY 6 TKNFVSTTISKSPAPRPRIQPSFGENVGKGLLFVNLTPENVSQVTVYVYDEY 65
 DB 478 TCGFSTVTTVSEPVQAEIPKIVKNVVDVQKNG-LVRKRLTLDNKNSETAVFRYIDL 536
 QY 66 --GLGLRVNT-----ADASQSIIVQYVDEKGRKMLKHGAETVNTQUTFKA--LNYT 114
 DB 537 SPALFLFTDTSFGQNSGGLVALVAVDPSTGILLMHGGS-----EPRADNQLDRT 589
 QY 115 SGEKK-----ISPGIYNDQVWVG 132
 DB 590 TAMKQGTGSSIKPIIYSAIETG 612
 RESULT 8
 ID 081IA3_BACCR PRELIMINARY; PRT; 749 AA.
 AC 081IA3
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Formate acetyltransferase (EC 2.3.1.54).
 GN OrderedLocustName=BC0491;
 OS *Bacillus cereus* (strain ATCC 14579 / DSM 31).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
 CC *Bacillus cereus* group.
 ON NCBI_TaxID=226900;
 RX [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
 RA Ivanova N., Sorokin A., Anderson I., Gallen N., Candelon B.,
 RA Kapatarel V., Bhattacharya A., Resnik G., Mkhallouf N., Lapidus A.,
 RA Chu L., Mazur M., Goetsman E., Larsen N., D'Souza M., Walunas T.,
 RA Grechkin Y., Pusch G., Haselkorn R., Fontein M., Ehrlich S.D.,
 RA Overbeek R., Kyriides N.C.;
 "Genome sequence of *Bacillus cereus* and comparative analysis with
 RT *Bacillus anthracis*.";
 RT Nature 423:87-91(2003).
 RL EMBL; AE016999; AAP07529.1; -; Genomic DNA.
 DR HSP; P09373; 1H16.
 DR GO; GO:0005737; C:cyttoplasm; IEA.
 DR GO; GO:0008415; F:acetyltransferase activity; IEA.
 DR GO; GO:0008861; F:formate C-acetyltransferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0006006; P:glucose metabolism; IEA.
 DR InterPro; IPR005949; Form_actrans.
 DR InterPro; IPR001150; Form_actrans.GR.
 DR InterPro; IPR004184; Pyr_Form_lyase.
 DR Pfam; PF01228; Gly_radical; 1.
 DR Pfam; PF02901; PF1.
 DR TIGRFAMs; TIGR01255; Pyr_form_ly_1; 1.
 DR PROSITE; PS00850; GLY_RADICAL; 1.

KW ACyltransferase; Complete proteome; Transferase.
 SO SEQUENCE 749 AA; 84681 MW; 5ZESAB51D0661002 CRC64;
 Query Match 11.9%; Score 83; DB 2; Length 749;
 Best Local Similarity 22.4%; Pred. No. 81;
 Matches 35; Conservative 28; Mismatches 67; Indels 26; Gaps 5;
 QY 4 IATN-PPVSTTISKSPAPRPRIQPSFGENVG--KGLLFVNLTPENVSQVTVYV 59
 DB 344 LVTKNSRFLTLTDLNLPAPRPNLTLVMSKQLPENFKYCARMSIKTSALQYENDDIRP 403
 QY 60 VYDEYGLGLRVNT-----ADASQSIIVQYVDEKGRKMLKHGAETVNTQUTFKA--LNYT 106
 DB 404 EYGDYGIACCVSAMRIGKMQPFGARANLAKALLIYALNGKDKSKRQVPEYPA---I 460
 QY 107 TFKALNTSGEKKIS-----PGIYNDQVWVYV 136
 DB 461 TSEVLNTEVNVHKFDMTWEMLAGLYLNTLVNTHYMH 496
 RESULT 9
 ID 08MQ08_CAEEL PRELIMINARY; PRT; 12268 AA.
 AC 08MQ08
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein K07E12.1.
 GN ORFNames=K07E12.1;
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 ON NCBI_TaxID=6239;
 RX [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=Bristol NZ;
 RC MEDLINE=99069613; PubMed=9851916;
 RX The C. elegans sequencing consortium;
 RT "Investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; U00054; AAM48546.1; -; Genomic DNA.
 DR HSP; P01130; 1H28.
 DR Ensembl; K07E12.1; *Caenorhabditis elegans*.
 DR WormBase; WBGene0019500; K07E12.1.
 DR WormPeP; K07E12.1b; CE31032.
 DR InterPro; IPR006209; EGF_1like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig_1like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR008162; Pyrophosphatase.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00008; EGF_1.
 DR Pfam; PF00041; FN3_10.
 DR Pfam; PF00047; Ig_3.
 DR Pfam; PF00084; Sushi_1.
 DR Pfam; PF00092; VWA_1.
 DR SMART; SM00032; CCP_1.
 DR SMART; SM00181; EGF_1.
 DR SMART; SM00060; FN3_10.
 DR SMART; SM00408; IGF2_1.
 DR SMART; SM00327; VWA_2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00853; FN3_10.
 DR PROSITE; PS00835; IG_1like; 7.
 DR PROSITE; PS00387; PRASE; UNKNOWN_1.
 DR PROSITE; PS0234; VWF_A; 2.
 KW Complete proteome; Hypothetical protein.
 SO SEQUENCE 12268 AA; 1282637 MW; 4418C6C048E635A6 CRC64;
 Query Match 11.9%; Score 83; DB 2; Length 12268;

```

DR PROSITE; PS01166; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS50853; FN3; 11.
DR PROSITE; PS50835; IG_LIKE; 7.
DR PROSITE; PS00387; PPASE; UNKNOWN_1.
DR PROSITE; PS50234; VMFA; 4.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 13100 AA; 1373302 MW; E13D0BF28C7EACFE CRC64;

Query Match 11.9%; Score 83; DB 2; Length 13100;
Best Local Similarity 25.6%; Pred. No. 27;
Matches 34; Conservative 23; Mismatches 50; Indels 28; Gaps 6;

Qy 10 PVSTTISKSPFAAPBPR--QPSFGENVKGGALLFSVNLVTPENVQSVTVYVYDEDEYGL 67
Db 10446 PLGDDSSGSYITDQGLVGDEEKGKPVGPQVL-----PTDSAGHYVPI---TGA 10499
Qy 68 GRLVNTVDAQSOSIIYQVDEDEKGGKMLKD-----HGAETVTPNQITFKALNTSGEKI 120
Db 10495 DRQILITDAGKRPITYSFNEDGIQLPTDSSGYAIGHGELVPTSTNGVPLN-----KDG 1054

Qy 121 SPGIYNDQVNVYV 135
Db 10550 TPLPTNDS---GHFV 10561

RESULT 11
Q6H0E5_BACAN PRELIMINARY; PRT; 248 AA.
AC 06H0E5_BACAN
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Methionine aminopeptidase (MAP) (Peptidase M).
OS OrderedLocustNames=BAS5204;
GN Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
OX NCBI_TaxID=1392;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Bredin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus anthracis Sterne.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL: A6017325; AAT57493.1; -; Genomic DNA.
DR GO:0004339; F:methionyl aminopeptidase activity; IEA.
DR GO:0008233; F:peptidase activity; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR001714; Pept_M24_MAP1.
DR InterPro: IPR002467; Pept_M24_MAP1.
DR InterPro: IPR000994; Peptidase_M24.
DR Pfam: PF00557; Peptidase_M24; 1.
DR PRINTS: PR00599; MAPPEPTIDASE.
DR TIGRFAMs: TIGR00500; met_hydase_I.1.
KW Aminopeptidase; Cobalt; Hydrolase; Protease.
SQ SEQUENCE 248 AA; 27062 MW; 14A6A817CE3BFD9E CRC64;

Query Match 11.7%; Score 82; DB 2; Length 248;
Best Local Similarity 25.6%; Pred. No. 27;
Matches 23; Conservative 18; Mismatches 29; Indels 20; Gaps 2;

Qy 67 IGRLVNTA-----DASOSIIYQVDEKGGKMLKDHAETVTPNQITFKALNTSGEKI 120
Db 15 IGRIVALAREMKKEAKPGKTTKELDLIGKIVLDEKHAISAPEKEDYFPVGVTSVNEEV 74

Qy 121 SPGIYNDQVNVYV-----GIYV 136
Db 75 AHGIPGDRVLKEGDVLNVVDVSAALDGYAD 104

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:13:10 ; Search time 189 Seconds
(without alignment)

316.167 Million cell updates/sec

Title: US-10-768-093-9

Sequence: 1 RTIATKMPVSTTISKFF.....EKISPCYNDQWVGYYVN 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*
- 9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	699	100.0	136	ADRI4916	Adri4916 CS6 relat
2	699	100.0	136	ADM47737	Adm47737 E. coli C
3	699	100.0	154	ADRI4912	Adri4912 Escherich
4	699	100.0	154	ADM47733	Adm47733 E. coli C
5	113.5	16.2	146	ADRI4917	Adri4917 CS6 relat
6	113.5	16.2	167	ADRI4913	Adri4913 Escherich
7	113.5	16.2	167	ADM47734	Adm47734 E. coli C
8	84.5	12.1	1239	ADRI6415	Adri6415 Aspergill
9	81.5	11.7	610	AE817679	Aeb17679 L. pneumo
10	81.5	11.7	610	AE817679	Aeb17679 L. pneumo
11	79.5	11.4	835	ABP26319	Abp26319 Streptoco
12	78.5	11.2	332	AAAG10781	Aag10781 Arabidops
13	78.5	11.2	335	AAAG10780	Aag10780 Arabidops
14	78.5	11.2	355	AEA36463	Aea36463 A. thalia
15	78.5	11.2	367	AAAG10779	Aag10779 Arabidops
16	78.5	11.2	378	ADCE9115	Adce9115 E. faeciu
17	78.5	11.2	268	ADM47927	Adm47927 Thermobif
18	77.5	11.1	754	ABM72032	Abm72032 Staphyloc
19	77	11.0	356	AEBA40460	Aeb40460 L. pneumo
20	76.5	10.9	384	AE837118	Aeb37118 L. pneumo
21	76.5	10.9	186	ADF06670	Adf06670 Bacterial
22	76	10.9	557	AD642954	Ad642954 Bacterial
23	76	10.9	2139	ADH62803	Adh62803 Lactobact
24	76	10.9	2358	ABU40683	Abu40683 Protein e

25	75.5	10.8	246	6	ABU39391
26	75.5	10.8	358	9	AEA36477
27	75.5	10.8	361	5	ABB93393
28	75.5	10.8	603	9	ABM92198
29	75	10.7	130	7	ADH67135
30	75	10.7	640	6	ABU16767
31	75	10.7	736	6	ABU35700
32	74.5	10.7	736	6	ABU30700
33	74.5	10.7	400	4	AA682840
34	74.5	10.7	497	5	ABP40269
35	74.5	10.7	497	5	ADSO6735
36	74.5	10.7	750	6	ADB99635
37	73.5	10.5	357	6	ABU39054
38	73.5	10.5	556	6	ABU50258
39	73.5	10.5	580	8	ADN46850
40	73	10.4	736	4	AAU35880
41	72.5	10.4	375	6	ADA89505
42	72.5	10.4	645	8	ADY23668
43	72.5	10.4	1773	9	ABB63908
44	72.5	10.4	2541	9	ABE96985
45	72	10.3	218	6	ADA35599

ALIGNMENTS

RESULT 1
ID ADRI4916 standard; protein; 136 AA.
AC ADRI4916;
DT 04-NOV-2004 (first entry)
DE CS6 related protein #1.
KW CS6; cs6A; cs6B; cs6C; cs6D; origin of replication; lac promoter;
KM kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.
OS Unidentified.
PN US004156829-A1.
PD 12-AUG-2004.
XX 02-FEB-2004; 2004US-00768093.
XX 13-MAY-1994; 94US-00243482.
PR 24-JAN-1997; 97US-00788145.
PR 10-JAN-2000; 2000US-00479877.
XX (US9A) US SEC OF ARMY.
PI Wolf MK, Casseels FU, Boedeker EC;
XX WPI; 2004-592722/57.
DR WPI; 2004-592722/57.
PT Inducing the production of antibodies against CS6 protein, useful for
PT preventing pathological effects of enterotoxigenic E. coli, by
PT administering a composition comprising transformed bacteria producing CS6
PT antigens.
PS Claim 4; SEQ ID NO 9; 22pp; English.
XX The invention relates to a method of inducing, in a susceptible host, the
XX production of antibodies against a CS6 protein, comprising administering
XX a composition of matter comprising bacteria transformed with a plasmid
XX containing genes cs6A, cs6B, cs6C and cs6D, an origin of replication, a
XX lac promoter and a kanamycin resistance gene, where the bacteria
XX expresses both CS6A and CS6B proteins. The invention also relates to a
XX composition of matter comprising a protein in a pharmaceutical carrier.
XX The pharmaceutical carrier is a carbonated liquid. The method is useful
XX for inducing the production of antibodies against the CS6 protein. The

protein and composition are used as vaccines for preventing pathological effects of enterotoxigenic E. coli. This sequence represents a CS6 related protein of the invention.

Sequence 136 AA;

Query Match 100.0%; Score 699; DB 8; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RTEATKMPVSTTTSKSPFAPEPRIOPSFGENVGKGLLFSVNLTVPENVSQVTVPV 60
DB 1 RTEATKMPVSTTTSKSPFAPEPRIOPSFGENVGKGLLFSVNLTVPENVSQVTVPV 60
OY YDEBYGLRLVNTADASQSIYQIVDEKGMKLDHGAELVPNQOITFKALNTSGEKKI 120
DB 61 YDEBYGLRLVNTADASQSIYQIVDEKGMKLDHGAELVPNQOITFKALNTSGEKKI 120
OY 121 SPGIYNDQVMVGYVN 136
DB 121 SPGIYNDQVMVGYVN 136

RESULT 2

ADM47737 standard; protein; 136 AA.

ADM47737;

21-APR-2005 (first entry)

E. coli CS6 csa mature protein.

antibacterial; vaccine; DNA purification; colonization factor antigen.

Escherichia coli.

US2005025787-A1.

03-FEB-2005.

12-JAN-2004; 2004US-00754641.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(WOLF/) WOLF M. K.

(CASS/) CASSELS F. J.

(BOED/) BOEDEKER E. C.

Wolf MK, Cassele FJ, Boedecker EC;

WPI; 2005-131784/14.

Inducing in a susceptible host, the production of antibodies against CS6

protein, useful for stimulating protective antibodies against Escherichia

coli by administering a composition comprising bacteria transformed with

a plasmid.

Claim 4; SEQ ID NO 9; 23pp; English.

The invention relates to a method of inducing in a susceptible host, the production of antibodies against CS6 protein comprising giving a composition of matter made of bacteria transformed with a plasmid, which contain genes csaA and csaB, all of csaC and DNA sequence csaD that encodes at least 802 amino acids (at least 2406 base pairs), an origin of replication, a lac promoter, and a kanamycin resistance gene, where the bacteria overexpress both CS6A and CS6B proteins. The method and proteins are useful for stimulating protective antibodies against enterotoxigenic Escherichia coli. This sequence corresponds to the E. coli CS6 csaA mature protein

Sequence 136 AA;

Query Match 100.0%; Score 699; DB 9; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.1e-72;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RTEATKMPVSTTTSKSPFAPEPRIOPSFGENVGKGLLFSVNLTVPENVSQVTVPV 60
DB 1 RTEATKMPVSTTTSKSPFAPEPRIOPSFGENVGKGLLFSVNLTVPENVSQVTVPV 60
OY YDEBYGLRLVNTADASQSIYQIVDEKGMKLDHGAELVPNQOITFKALNTSGEKKI 120
DB 61 YDEBYGLRLVNTADASQSIYQIVDEKGMKLDHGAELVPNQOITFKALNTSGEKKI 120
OY 121 SPGIYNDQVMVGYVN 136
DB 121 SPGIYNDQVMVGYVN 136

RESULT 3

ADR14912 standard; protein; 154 AA.

ADR14912;

04-NOV-2004 (first entry)

Escherichia coli E8775 antigen csaA protein.

CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter;

kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli;

Escherichia coli E8775.

Escherichia coli.

US2004156829-A1.

12-AUG-2004.

02-FEB-2004; 2004US-00768093.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(USSA) US SEC OF ARMY.

Wolf MK, Cassele FJ, Boedecker EC;

WPI; 2004-592722/57.

N-PSDB; ADR14911.

Inducing the production of antibodies against CS6 protein, useful for

preventing pathological effects of enterotoxigenic E. coli, by

administering a composition comprising transformed bacteria producing CS6

antigens.

Disclosure; SEQ ID NO 5; 22pp; English.

The invention relates to a method of inducing, in a susceptible host, the production of antibodies against a CS6 protein, comprising administering a composition of matter comprising bacteria transformed with a plasmid containing genes csaA, csaB, csaC and csaD, an origin of replication, a lac promoter and a kanamycin resistance gene, where the bacteria express both CS6A and CS6B proteins. The invention also relates to a composition of matter comprising a protein in a pharmaceutical carrier. The pharmaceutical carrier is a carbonated liquid. The method is useful for inducing the production of antibodies against the CS6 protein. The protein and composition are used as vaccines for preventing pathological effects of enterotoxigenic E. coli. This sequence represents the Escherichia coli E8775 antigen csaA of the invention.

Sequence 154 AA;

Query Match 100.0%; Score 699; DB 8; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.3e-72;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNFVSTTISKSFPAPRIQPSFGENVKKGALLFSVNLTPENVSGVTYYPV 60
 DB 19 RTEIATKNFVSTTISKSFPAPRIQPSFGENVKKGALLFSVNLTPENVSGVTYYPV 78
 QY 61 YDEBYGIGRLVNTADASQSIITQIVDEKKGKMLKDHAETVPNOQITFKALNTSGEKI 120
 DB 79 YDEBYGIGRLVNTADASQSIITQIVDEKKGKMLKDHAETVPNOQITFKALNTSGEKI 138
 QY 121 SPGIYNDQVMVGYVYN 136
 DB 139 SPGIYNDQVMVGYVYN 154

RESULT 4

ID ADM47733 standard; protein; 154 AA.

AC ADM47733;
 DT 21-APR-2005 (first entry)
 DE E. coli CS6 csaA full length protein.
 KM antibacterial; vaccine; DNA purification; colonization factor antigen.
 OS Escherichia coli.

PH Key Location/Qualifiers
 FT Peptide 1..18
 FT Protein /note= "signal peptide"
 FT /note= "mature peptide"

US2005025787-A1.

03-FEB-2005.

12-JAN-2004; 2004US-00754641.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(WOLF/) WOLF M K.

(CASS/) CASSELS F J.

(BOED/) BOEDEKER E C.

WOLF MK, Casseels FJ, Boedecker EC;

WPI; 2005-131784/14.

N-PSDB; ADM47732.

Inducing in a susceptible host, the production of antibodies against CS6

protein, useful for stimulating protective antibodies against Escherichia

coli by administering a composition comprising bacteria transformed with

a plasmid.

Disclosure; SEQ ID NO 5; 23pp; English.

Applicant

CC protein
 XX Sequence 154 AA;
 SQ

Query Match 100.0%; Score 699; DB 9; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.3e-72;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNFVSTTISKSFPAPRIQPSFGENVKKGALLFSVNLTPENVSGVTYYPV 60
 DB 19 RTEIATKNFVSTTISKSFPAPRIQPSFGENVKKGALLFSVNLTPENVSGVTYYPV 78
 QY 61 YDEBYGIGRLVNTADASQSIITQIVDEKKGKMLKDHAETVPNOQITFKALNTSGEKI 120
 DB 79 YDEBYGIGRLVNTADASQSIITQIVDEKKGKMLKDHAETVPNOQITFKALNTSGEKI 138
 QY 121 SPGIYNDQVMVGYVYN 136
 DB 139 SPGIYNDQVMVGYVYN 154

RESULT 5

ID ADR14917 standard; protein; 146 AA.

AC ADR14917;

DT 04-NOV-2004 (first entry)

DE CS6 related protein #2.

KM CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter;

KW kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.

OS Unidentified.

US2004156829-A1.

12-AUG-2004.

02-FEB-2004; 2004US-00768093.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(USSA) US SEC OF ARMY.

WOLF MK, Casseels FJ, Boedecker EC;

WPI; 2004-592722/57.

Inducing the production of antibodies against CS6 protein, useful for

preventing pathological effects of enterotoxigenic E. coli, by

administering a composition comprising bacteria transformed bacteria producing CS6

antigens.

Claim 6; SEQ ID NO 10; 22pp; English.

The invention relates to a method of inducing, in a susceptible host, the

production of antibodies against a CS6 protein, comprising administering

a composition of matter comprising bacteria transformed with a plasmid

containing genes csaA, csaB, csaC and csaD, an origin of replication, a

lac promoter and a kanamycin resistance gene, where the bacteria

expresses both CS6A and CS6B proteins. The invention also relates to a

composition of matter comprising a protein in a pharmaceutical carrier.

The pharmaceutical carrier is a carbonated liquid. The method is useful

for inducing the production of antibodies against the CS6 protein. The

protein and composition are used as vaccines for preventing pathological

effects of enterotoxigenic E. coli. This sequence represents a CS6

related protein of the invention.

Sequence 146 AA;

Applicant

Query Match 16.2%; Score 113.5; DB 8; Length 146;
Best Local Similarity 28.5%; Pred. No. 0.00014;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVKKGALLFSVNLTVPENVSQVTVYP--- 59
DB 6 KSLDVNVNIQNFI-PDIDSAVRILIPVNYDSPKLDQSLYVEMTIPAGVSAVKIAPIDS 64
OY 60 VYDEBYGRLVNTADASQSIITYQIVDEKGMKLDHGA-----EYTPNQQTTF 108
DB 65 LTSSGQOIGKLVNVPDQNNYYI-----RKDSGAGNFMAGQKGFVKENTSYTF 116
OY 109 KALNTSGE---KKISPGIYNDQVWVGYVN 136
DB 117 SAI-VTGGEPNPGYSSTGTAGNLTVSFYSN 146

RESULT 6
ID ADR14913 standard; protein; 167 AA.
XX ADR14913;
AC ADR14913;
DT 04-NOV-2004 (first entry)
XX
XX Escherichia coli E8775 antigen cs6B protein.
DE
XX
XX CS6; cs6A, cs6B; cs6C; cs6D; origin of replication; lac promoter;
KW kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli;
KW Escherichia coli E8775.
XX
XX Escherichia coli.
OS
XX US2004156829-A1.
XX
XX 12-AUG-2004.
PD
XX 02-FEB-2004; 2004US-00768093.
PF
XX 13-MAY-1994; 94US-00243482.
PR 24-JAN-1997; 97US-00788145.
PR 10-JAN-2000; 2000US-00479877.
XX
XX (USSA) US SEC OF ARMY.
PA
XX
XX Wolf MK, Casseels FJ, Boedeker EC;
PI
XX
XX WPI; 2004-592722/57.
DR
XX
XX Inducing the production of antibodies against CS6 protein, useful for
PT preventing pathological effects of enterotoxigenic E. coli, by
PT administering a composition comprising transformed bacteria producing CS6
PT antigens.
PS
XX
XX Disclosure; SEQ ID NO 6; 22pp; English.
XX
XX The invention relates to a method of inducing, in a susceptible host, the
CC production of antibodies against a CS6 protein, comprising administering
CC a composition of matter comprising bacteria transformed with a plasmid
CC containing genes cs6A, cs6B, cs6C and cs6D, an origin of replication, a
CC lac promoter and a kanamycin resistance gene, where the bacteria
CC express both CS6A and CS6B proteins. The invention also relates to a
CC composition of matter comprising a protein in a pharmaceutical carrier.
CC The pharmaceutical carrier is a carbonated liquid. The method is useful
CC for inducing the production of antibodies against the CS6 protein. The
CC protein and composition are used as vaccines for preventing pathological
CC effects of enterotoxigenic E. coli. This sequence represents the
CC Escherichia coli E8775 antigen cs6B of the invention.
XX
XX Sequence 167 AA;
SQ

Query Match 16.2%; Score 113.5; DB 8; Length 167;

Best Local Similarity 28.5%; Pred. No. 0.00017;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVKKGALLFSVNLTVPENVSQVTVYP--- 59
DB 27 KSLDVNVNIQNFI-PDIDSAVRILIPVNYDSPKLDQSLYVEMTIPAGVSAVKIAPIDS 85
OY 60 VYDEBYGRLVNTADASQSIITYQIVDEKGMKLDHGA-----EYTPNQQTTF 108
DB 86 LTSSGQOIGKLVNVPDQNNYYI-----RKDSGAGNFMAGQKGFVKENTSYTF 137
OY 109 KALNTSGE---KKISPGIYNDQVWVGYVN 136
DB 138 SAI-VTGGEPNPGYSSTGTAGNLTVSFYSN 167

RESULT 7
ID ADM47734 standard; protein; 167 AA.
XX ADM47734;
AC ADM47734;
DT 21-APR-2005 (first entry)
XX
XX E. coli CS6 cs6B full length protein.
DE
XX
XX antibacterial; vaccine; DNA purification; colonization factor antigen.
KW
XX
XX Escherichia coli.
OS
XX US2005025787-A1.
PN
XX
XX 03-FEB-2005.
PD
XX 12-JAN-2004; 2004US-00754641.
PF
XX 13-MAY-1994; 94US-00243482.
PR 24-JAN-1997; 97US-00788145.
PR 10-JAN-2000; 2000US-00479877.
XX
XX (WOLF/) WOLF M K.
PA (CASS/) CASSELS F J.
PA (BOED/) BOEDEKER E C.
XX
XX
XX Wolf MK, Casseels FJ, Boedeker EC;
PI
XX
XX WPI; 2005-131784/14.
DR
XX
XX Inducing in a susceptible host, the production of antibodies against CS6
PT protein, useful for stimulating protective antibodies against Escherichia
PT coli by administering a composition comprising bacteria transformed with
PT a plasmid.
PS
XX
XX Disclosure; SEQ ID NO 6; 23pp; English.
XX
XX The invention relates to a method of inducing in a susceptible host, the
CC production of antibodies against CS6 protein comprising giving a
CC composition of matter made of bacteria transformed with a plasmid, which
CC contain genes cs6A and cs6B, all of cs6C and DNA sequence cs6D that
CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of
CC replication, a lac promoter, and a kanamycin resistance gene, where the
CC bacteria overexpress both CS6A and CS6B proteins. The method and proteins
CC are useful for stimulating protective antibodies against enterotoxigenic
CC Escherichia coli. This sequence corresponds to the E. coli CS6 cs6B
CC protein.
XX
XX Sequence 167 AA;
SQ

Query Match 16.2%; Score 113.5; DB 9; Length 167;
Best Local Similarity 28.5%; Pred. No. 0.00017;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVKKGALLFSVNLTVPENVSQVTVYP--- 59

```

Db      27 KSLDVLNVNIEQNF1-PIDSAVR1IPVNYDSDPKLDLSQLYTVENTITPAGVSAVXIAPTDS 85
Qy      60 VYDEBDYIGRLVNTADASQSI-YQIVDEKGMKMDHGA-----EVTNQGITP 108
Db      86 LTSSGQOIGKLVNVPDQNNYTI-----RKDSAGNFMAGOKSPFKENTSYTF 137
Qy      109 KALNVTSGE---KKISPGIYNDQWVGYYVN 136
Db      138 SAI-YTGGEYPNSGSYSGTYAGNLTVSFYSN 167

RESULT 8
ADR86415
ID      ADR86415 standard; protein; 1239 AA.
XX
AC      ADR86415;
XX
DT      04-NOV-2004 (first entry)
XX
DE      Aspergillus fumigatus essential gene protein #465.
XX
KW      Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
XX      drug screening.
XX
OS      Aspergillus fumigatus.
XX
PN      WO2004067709-A2.
XX
PD      12-AUG-2004.
XX
PF      16-JAN-2004; 2004WO-US001099.
XX
PR      17-JAN-2003; 2003US-0441281P.
XX
PR      13-JUN-2003; 2003US-0478196P.
XX
PA      (ELIT-) ELITRA PHARM INC.
XX      (ELIT-) ELITRA CANADA LTD.
XX
PI      Jiang B, Hu W, Lemieux S, Roemer T;
XX
DR      WPI; 2004-594200/57.
XX
DR      N-PSDB; ADR85828.
XX
PT      New purified or isolated Aspergillus fumigatus nucleic acid molecule
XX      encoding a gene product, useful for diagnosing and/or treating invasive
XX      fungal infections, such as Farmer's lung disease.
XX
PS      Claim 1; SEQ ID NO 3465; 164bp; English.
XX
CC      The present invention relates to Aspergillus fumigatus genes that are
XX      essential and are potential targets for drug screening. The methods and
XX      compositions of the present invention are useful for diagnosing and/or
XX      treating invasive Aspergillus fumigatus infection. Including the allergic
XX      forms of the disease, such as Farmer's lung disease. They can also be
XX      used in various drug discovery purposes, such as expression of the
XX      recombinant protein, hybridization assay and construction of nucleic acid
XX      arrays. The present sequence represents an Aspergillus fumigatus
XX      essential gene protein sequence, used during diagnosis and drug
XX      development in the invention. These genes share a high degree of sequence
XX      conservation with known essential genes of candida albicans. The sequence
XX      data for this patent is not represented in the printed specification, but
XX      was obtained in electronic format from WIPO.
XX
SQ      Sequence 1239 AA;
XX
Query Match      12.1%; Score 84.5; DB 8; Length 1239;
Best Local Similarity 28.0%; Pred. No. 8;
Matches 35; Conservative 20; Mismatches 53; Indels 17; Gaps 6;

Qy      12 STTISKFPFAR-----EPRIQSPFGENVGKSGALLFSVNLVTPENVQVTVYPYVDE 63
Db      1031 ATTIVNVQSPFAANGTKAPVSEP---EGTAKERSARTLGLMNVPTVDARIRAMV-E 1086

```

```

Qy      64 DYGLGELVNTADASQSI-YQIVDEKGMKMDHGEVTPNQI---TFKALNVTSGEK 118
Db      1087 PYSPFLIKIVLPHQGIYVFAVDNHNAGKASLEEGGEIAPGQRLHVGTVSEILKQSAEK 1146
Qy      119 KISPG 123
Db      1147 KGGPG 1151

RESULT 9
AEB37679
ID      AEB37679 standard; protein; 610 AA.
XX
AC      AEB37679;
XX
DT      08-SEP-2005 (first entry)
XX
DE      L. pneumophila protein SEQ ID NO 2011.
XX
KW      detection; infection; Antibacterial; Vaccine.
XX
OS      Legionella pneumophila.
XX
PN      WO2005049642-A2.
XX
PD      02-JUN-2005.
XX
PF      23-SEP-2004; 2004WO-IB003578.
XX
PR      21-NOV-2003; 2003FR-00013687.
XX
PA      (INSP ) INST PASTEUR.
XX      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX      (UPLY-) UNIV LYON 1 BERNARD CLAUDE.
XX      (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI      Buchrieser C, Tichit M, Etienne J, Ma L, Cazalec C, Glaser P;
XX      Rusnlok C, Bouchier C, Zidane N, Magnier A, Kunet F, Vandenesch F;
XX      Jarraud S;
XX
DR      WPI; 2005-388305/40.
XX
PT      New genome of Legionella pneumophila Paris strain and derived
XX      polypeptides, useful for detection or identification of the strain and
XX      for treatment and prevention of infections.
XX
PS      Claim 3; SEQ ID NO 2011; 660bp; English.
XX
CC      The invention relates to an isolated or purified nucleotide sequences (I)
XX      from Legionella pneumophila Paris strain. (I), and their related
XX      sequences or fragments, are useful as primers and probes for detection
XX      and amplification, including differentiation between the Paris and
XX      Philadelphia strains of Legionella pneumophila and to prepare recombinant
XX      (hybrid) polypeptides (II). (II) are also useful for preparation of
XX      specific antibodies (Ab), also used for detection/identification of
XX      Legionella, and some (I), specifically those involved in synthesis of
XX      surface proteins, are targets for identification of inhibitors. (II), or
XX      vectors that contain (I), are useful as vaccines and immunogenic
XX      compositions, for treatment and prevention of infections by L.
XX      pneumophila. The present sequence represents the amino acid sequence of a
XX      L. pneumophila protein.
XX
SQ      Sequence 610 AA;
XX
Query Match      11.7%; Score 81.5; DB 9; Length 610;
Best Local Similarity 25.7%; Pred. No. 6.3;
Matches 29; Conservative 19; Mismatches 34; Indels 31; Gaps 5;

Qy      9 FPVST-----TISKSPFAPRIQSPFGENVGKSGALLFSVNLTPENV 53
Db      312 FPIVSSDDEFAFRALKLSTNDASLFPYEPESSEALGPGFCFLGMLHMEI---IQRL- 367

```

QY 54 QVTVPYVDEYDGLGRLVNTADASQSIIVQIVDEKXKMLKDGAETVPENOOI 106
DB 368 -----ERENVLD-LISTA---PTVVYQIVTQGETLLIDNPSHLPTPTQI 408

RESULT 10
ID AEB40990 standard; protein; 610 AA.
AC AEB40990;
XX
XX 08-SEP-2005 (first entry)
XX
XX L. pneumophila protein SEQ ID NO 5322.
XX
XX detection; infection; Antibacterial; Vaccine.
XX
XX Legionella pneumophila.
XX
XX WO2005049642-A2.
XX
XX 02-JUN-2005.
XX
XX 23-SEP-2004; 2004MO-IB003578.
XX
XX 21-NOV-2003; 2003FR-00013687.
XX
XX (INSP) INST PASTEUR.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX (UTLY-) UNIV LYON 1 BERNARD CLAUDE.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Buchrieser C, Tichit M, Etienne J, Ma L, Cazalec C, Glaser P,
PI Ruenick C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F,
PI Jarraud S;
XX WPI: 2005-388305/40.
XX
XX New genome of Legionella pneumophila Paris strain and derived
PT polypeptides, useful for detection or identification of the strain and
PT for treatment and prevention of infections.
XX
XX Claim 3; SEQ ID NO 5322; 660bp; English.
XX
XX The invention relates to an isolated or purified nucleotide sequences (I)
CC from Legionella pneumophila Paris strain. (I), and their related
CC sequences or fragments, are useful as primers and probes for detection
CC and amplification, including differentiation between the Paris and
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant
CC (hybrid) polypeptides (II). (II) are also useful for preparation of
CC specific antibodies (Ab), also used for detection/identification of
CC Legionella, and some (I), specifically those involved in synthesis of
CC surface proteins, are targets for identification of inhibitors. (II), or
CC vectors that contain (I), are useful as vaccines and immunogenic
CC compositions, for treatment and prevention of infections by L.
CC pneumophila. The present sequence represents the amino acid sequence of a
CC L. pneumophila protein.
XX
XX
SQ Sequence 610 AA;
Query Match 11.7%; Score 81.5; DB 9; Length 610;
Best Local Similarity 25.7%; Pred. No. 6.3;
Matches 29; Conservative 19; Mismatches 34; Indels 31; Gaps 5;

QY 9 FTVST-----TISKSPFAPRIPGPGNVKKGALLFSVVLTVPEVNS 53
DB 312 FVVSSDDPEAFBEALAKSLNDASLFTPESESEALGFRGCGFLMLMEI---IQEHL- 367

QY 54 QVTVPYVDEYDGLGRLVNTADASQSIIVQIVDEKXKMLKDGAETVPENOOI 106
DB 368 -----ERENVLD-LISTA---PTVVYQIVTQGETLLIDNPSHLPTPTQI 408

RESULT 11
ID ABP26319 standard; protein; 835 AA.
AC ABP26319;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 1814.
XX
XX Streptococcus pyogenes; GAS; group A streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001MO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX 24-NOV-2000; 2000GB-00028727.
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C,
PI Tettein H;
XX WPI: 2002-352536/38.
XX N-PSDB; ABN66950.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3331; 4525bp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX
SQ Sequence 835 AA;
Query Match 11.4%; Score 79.5; DB 5; Length 835;
Best Local Similarity 21.8%; Pred. No. 1.7;
Matches 29; Conservative 28; Mismatches 31; Indels 45; Gaps 7;

QY 13 TTISKSPFAPRIPGPGNVKKGALLFSVVLTVPEVNS-QVTVPYVDEYDGLGRLV 71
DB 412 TPTLKTTPRQPKRP-----HVSISEKINYSVHPV-----LV 446

QY 72 NTADASQSIIVQIVDEKXKMLKDGAETVPENOOITPKA-----LNTSGEKKISFGI 124
DB 447 PAAKPSKAVI---DEAGQSV---NGKTVLFPNALNLVAKODFQYQKMTSOKIAXNF 499

QY 125 -----YNDQWVG 132
Db 500 VFIDYKDDALDG 512

RESULT 12
AAG10781
ID AAG10781 standard; protein; 332 AA.
XX AAG10781,
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 9239.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131444P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 19-MAY-1999; 99US-0135124P.
PR 20-MAY-1999; 99US-0135353P.
PR 21-MAY-1999; 99US-0135629P.
PR 24-MAY-1999; 99US-0136021P.
PR 25-MAY-1999; 99US-0136392P.
PR 27-MAY-1999; 99US-0136782P.
PR 28-MAY-1999; 99US-0137222P.
PR 01-JUN-1999; 99US-0137528P.
PR 03-JUN-1999; 99US-0137529P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.

PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149928P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158025P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 13-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.2%; Score 78.5; DB 3; Length 332;

Best Local Similarity 27.3%; Pred. No. 5.7; Mismatches 39; Indels 29; Gaps 6;

Matches 33; Conservative 20; Mismatches 39; Indels 29; Gaps 6;

DB 10 PVSTTI--SKSFAPDPRIQPSFGENVGKALLFESVNLTVBPNSQVTVFYVDY-- 65
DB 193 PVGSIYVSKKFTYARWRKTLGGGMRQIGVLCALVALHENVAKL-----EDDHKK 246
DB 66 -----GICRL-----VNTDAQSITTYOVDKGR-----KMKDHGAETVTPNQITF 108
DB 247 ARVLAEGNIRLRNVAAVETNIIYVIDPEDPKFGAEACKSLDVGVLVIP--QATF 304

XX 109 K 109
DB 305 R 305
RESULT 13
AAgi0780
ID AAgi0780 standard; protein; 355 AA.
XX
AC AAgi0780;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9238.
XX
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN Epi033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134376P.
PR 18-MAY-1999; 99US-0134941P.
PR 19-MAY-1999; 99US-0135124P.
PR 20-MAY-1999; 99US-0135353P.
PR 21-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137502P.
PR 04-JUN-1999; 99US-0137724P.
PR 07-JUN-1999; 99US-0138094P.
PR 08-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.


```

PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145275P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147193P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148314P.
PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158223P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query March 11.2%; Score 78.5; DB 3; Length 355;
Best Local Similarity 27.3%; Pred. No. 6.3;
Matches 33; Conservative 20; Mismatches 39; Indels 29; Gaps 6;

```

DB 270 ARVLAEGLNRIERLRVNAVAETNIIIVYDIPEDPKFGAEACKSLDVGVLVIP--QATF 327
 OY 109 K 109
 DB 328 R 328

RESULT 14
 AEA36463
 ID AEA36463 standard; protein; 355 AA.
 AC AEA36463;
 XX
 AC 28-JUL-2005 (first entry)
 DT
 XX A. thaliana chromosome 3 threonine aldolase.
 DE
 XX threonine aldolase; silencing; mitochondrial targeting sequence;
 KM transgenic; soybean; alternative oxidase; potato; formate dehydrogenase;
 KM Nicotiana glauca; mitochondrial ATP synthase beta-subunit;
 KM animal feed; foodstuff.
 XX
 XX Arabidopsis thaliana.
 OS
 PN WO2005047472-A2.
 XX
 PD 26-MAY-2005.
 XX
 PF 12-NOV-2004; 2004WO-US037369.
 XX
 PR 10-NOV-2003; 2003US-0519313P.
 PA (BOYC-) BOYCE THOMPSON INST PLANT RES INC.
 PI Jander G, Joshi V;
 XX WPI; 2005-372364/38.
 DR N-PSDB; AEA36464.
 XX
 PT New nucleic acid construct comprising a nucleic acid molecule,
 PT mitochondrial targeting, promoter, and terminator sequences, useful for
 PT increasing threonine content in plant seeds.
 XX
 PS Claim 2; SEQ ID NO 3; 130bp; English.

This sequence represents a threonine aldolase. This threonine aldolase is an example of a threonine aldolase which may be silenced by the nucleic acid construct of the invention which comprises a nucleic acid molecule configured to silence threonine aldolase expression or encodes a threonine aldolase, a mitochondrial targeting sequence, a 5' DNA promoter sequence, and a 3' terminator sequence, where the nucleic acid molecule, the mitochondrial targeting sequence, the promoter, and the terminator are operatively coupled to permit transcription of the nucleic acid molecule. This nucleic acid construct may be used in the generation of transgenic plants which have increased levels of seed threonine and/or isoleucine. The mitochondrial targeting sequence comprises a mitochondrial targeting leader sequence, e.g. a soybean alternative oxidase gene, a potato formate dehydrogenase gene, or a Nicotiana glauca plasmid. The nucleic acid molecule may encode a nucleic acid molecule that may comprise a dominant negative mutation and encode a nonfunctional threonine aldolase, resulting in suppression or interference of endogenous mRNA encoding threonine aldolase. The nucleic acid molecule is positioned in the nucleic acid construct to result in suppression or interference of endogenous mRNA encoding threonine aldolase. The nucleic acid molecule may encode threonine aldolase and be in the same orientation, or the nucleic acid molecule is an antisense form or a threonine aldolase encoding nucleic acid molecule. The mutation may be caused by a chemical mutagenizing agent, e.g. ethylmethanesulphonate, radiation, e.g. ultraviolet rays, gamma rays, or fast neutrons, or by inserting an inactivating nucleic acid molecule into the gene encoding the functional threonine aldolase or its promoter under conditions to inactivate the gene, where the deactivating nucleic acid molecule is a transposable element, e.g. an Activator (Ac) transposon, a

CC Dissociator (Ds) Transposon, or a Mutator (Mu) transposon. Increasing
 CC isoleucine content in a plant comprises providing a non-wild-type plant
 CC having increased threonine content compared to a corresponding wild-type
 CC plant, transforming the non-wild-type plant with a nucleic acid construct
 CC described above under conditions to yield a transgenic plant that over
 CC expresses an enzyme that functions to catalyze biosynthesis of isoleucine
 CC from threonine, and growing the transgenic plant under conditions to
 CC increase the isoleucine content of the transgenic plant. The nucleic acid
 CC construct and methods are useful for increasing threonine content in
 CC plants, particularly in plant seeds, which are then used in producing
 CC animal feed and foodstuff.
 XX
 SQ Sequence 355 AA;

Query Match 11.2%; Score 78.5; DB 9; Length 355;
 Best Local Similarity 27.3%; Pred. No. 6.3;
 Matches 33; Conservative 20; Mismatches 39; Indels 29; Gaps 6;

OY 10 PUSTT--SKSFAPERIPSPFGENVKKGALLFSVLTVPENVSQTVVPVDEY-- 65
 DB 216 PVGSVTVSGKKFTYKARLRLTGGGKMQIGVLCNALVALHENVAKL-----EDDHKK 269
 OY 66 -----GKRL-----VNTADASQSIYYQIVDEKX-----KMLKDHGAETVPPNQITTF 108
 DB 270 ARVLAEGLNRIERLRVNAVAETNIIIVYDIPEDPKFGAEACKSLDVGVLVIP--QATF 327
 OY 109 K 109
 DB 328 R 328

RESULT 15
 AAG10779
 ID AAG10779 standard; protein; 367 AA.
 XX
 AC AAG10779;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 9237.
 XX
 KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0128845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131444P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134222P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138009P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 06-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143547P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147322P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147203P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148652P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152353P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158359P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159847P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:17:05 ; Search time 39 Seconds
(without alignments)
335.525 Million cell updates/sec

Title: US-10-768-093-9

Sequence: 1 RTETATKPFVSTTISKSF.....EKISPGIYNDQVWVGYVN 136

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 80: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	655	93.7	154 2	160266 CS6 structural sub
2	120.5	17.2	167 2	183348 CS6 structural sub
3	83	11.9	13055 2	T16580 hypothetical prote
4	81	11.6	595 2	AF2097 hypothetical prote
5	78	11.2	608 2	AB1293 probable flagellar
6	77.5	11.1	740 1	JC6010 formate C-acetylitr
7	77.5	11.1	749 2	B89785 hypothetical prote
8	77.5	11.1	4936 2	AH2515 hypothetical prote
9	77	11.0	356 2	T25354 hypothetical prote
10	76.5	10.9	569 2	T13531 probable potassiu
11	76	10.9	557 2	F69481 probable acid-CoA
12	75.5	10.8	361 2	T49882 pectin methyl-este
13	75	10.7	266 2	A82851 chaperone protein
14	75	10.7	736 2	DE4534 DNA topoisomerase
15	75	10.7	1472 2	H82802 fibrillar assembly
16	74.5	10.7	946 2	UC5375 inner-alpha-trypt
17	74	10.6	275 2	UC7604 CS66 spliced varia
18	73.5	10.5	521 2	S62794 probable lipoprote
19	73.5	10.5	556 2	A10134 probable phage pro
20	73.5	10.5	718 2	T47930 hypothetical prote
21	73.5	10.5	881 2	F69438 conserved hypotet
22	73	10.4	736 2	B11972 DNA topoisomerase
23	73	10.4	1487 2	AG2560 hypothetical prote
24	73	10.4	1843 2	S18803 collagen alpha 1(V
25	72.5	10.4	925 2	S50490 hypothetical prote
26	72	10.3	880 2	AF2128 hypothetical prote
27	71.5	10.2	357 2	S71772 naringenin 3-dioxy
28	71.5	10.2	357 2	S61415 naringenin 3-dioxy
29	71.5	10.2	744 2	F75625 serine proteinase,

30	71.5	10.2	812 2	T01618 hypothetical prote
31	71	10.2	318 2	TS0039 beta-1,4-galactosy
32	71	10.2	1157 2	F97255 fusion of alpha-gl
33	70.5	10.1	217 2	T29388 hypothetical prote
34	70.5	10.1	250 2	F72247 methionine aminope
35	70.5	10.1	323 2	S69647 hypothetical prote
36	70.5	10.1	357 1	G64053 chorismate synthas
37	70.5	10.1	430 2	AG1381 enolase homolog en
38	70.5	10.1	430 2	AH1750 thiophen and furan
39	70.5	10.1	458 2	F84157 dihydropterocate sy
40	70.5	10.1	498 2	C90413 conserved membrane
41	70.5	10.1	896 2	AB1156 probable lipoprote
42	70.5	10.1	1644 2	AC0823 toxin B - Clostrid
43	70.5	10.1	2366 2	S10317 hypothetical prote
44	70.5	10.1	2911 2	T20566 outer membrane pro
45	70	10.0	242 2	H82720

ALIGNMENTS

RESULT 1
160266 CS6 structural subunit A - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 09-Jul-2004
C:Accession: 160266
R:Willshaw, G.A.; Smith, H.R.; McConnell, M.M.; Rowe, B.
FEMS Microbiol. Lett. 49, 473-478, 1988
A:Title: Cloning of genes encoding coli-surface (CS) antigens in enterotoxigenic Escheri
A:Reference number: 160266
A:Accession: 160266
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-154 <RES>
A:Cross-references: UNIPROT:P53508; UNIPARC:UP1000012A400; EMBL:U04844; NID:9442375; PID

Query Match
Best Local Similarity 93.7% Score 655, DB 2; Length 154;
Pred. No. 1,4e-54;
Matches 125; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY	1	RTETATKPFVSTTISKSFAPRQPSYGENYKGEKALFFYNLWPNVSVVTPV 60
DB	19	RTETATKPFVSTTISKSFAPRQPSYGENYKGEKALFFYNLWPNVSVVTPV 78
QY	61	YDEBYGGLRLVNTADASIIYIVDEKGRKMLKHGAETVPOQITFQALNTYSGEKI 120
DB	79	YDEBYGGLRLVNTADASIIYIVDEKGRKMLKHGAETVPOQITFQALNTYSGEKI 138
QY	121	SPGIYNDQVWVGYVN 136
DB	138	SPGIYNDQVWVGYVN 154

RESULT 2
183348 CS6 structural subunit B - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 09-Jul-2004
C:Accession: 183348
R:Willshaw, G.A.; Smith, H.R.; McConnell, M.M.; Rowe, B.
FEMS Microbiol. Lett. 49, 473-478, 1988
A:Title: Cloning of genes encoding coli-surface (CS) antigens in enterotoxigenic Escheri
A:Reference number: 160266
A:Accession: 183348
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-167 <RES>
A:Cross-references: UNIPROT:P53510; UNIPARC:UP1000012A402; EMBL:U04844; NID:9442375; PID

Query Match
Best Local Similarity 30.1% Score 120.5, DB 2; Length 167;
Pred. No. 0.00041;
Matches 43; Conservative 24; Mismatches 61; Indels 15; Gaps 7;

```

Oy 7 KNFVSTTISKSFPAPB-----PRIOPSEGENVKEGALLFSVNLTVPENVSQVTVYP--- 59
Db 27 KSLDNNVNI EQNFI-PDIDSAVRILIPVNYDSDPKLNSQLYTEMTIPGASAVKLVPTDS 85
Oy 60 VYDEDEYGLVNTADASQSIYQIVDEK--GKKMLKDHGA-EVTPNQOITFKALNTYSG 116
Db 86 LTSSGOQIGKLVNPNPQNNMYIRKDSGAGKFMAGKGSFVSXKNTSYFSAL-YTGG 144
Oy 117 E---KKISPGIYNDQVMVGYVYN 136
Db 145 EYPSNGVSSGTIYAGHUTISFYISN 167

RESULT 3
116580
hypotheoretical protein K07E12.1 - Caenorhabditis elegans
C:/Species: Caenorhabditis elegans
C:/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:/Accession: T16580
R:/Fulton, L.
Submitted to the EMBL Data Library, May 1994
A:/Description: The sequence of C. elegans coemid K07E12.
A:/Reference number: Z18540
A:/Accession: T16580
A:/Status: preliminary; translated from GB/EMBL/DBJ
A:/Molecule type: DNA
A:/Residues: 1-13055 <FUL>
A:/Cross-references: UNIPROT:O09165; UNIPARC:UPI000017CF3A; EMBL:U00054; NID:g485140; PID
C:/Genetics:
A:/Experimental source: strain Bristol N2
C:/Gene: CESP:K07E12.1
A:/Intons: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3;
: 6014/3; 6159/3; 6665/2; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1229

Query Match 11.9%; Score 83; DB 2; Length 13055;
Best Local Similarity 25.2%; Pred. No. 3.5e+02;
Matches 34; Conservative 23; Mismatches 50; Indels 28; Gaps 6;

Oy 10 PVSTTISGSFPAPBPRF-----QPSFGENVKEGALLFSVNLTVPENVSQVTVYPVYDEYGL 67
Db 10401 PLGTDSGSGYTTEDGQLVKQKEGKFPVDPDGVL-----PTDSAGHYVYPI-----TGA 10449
Oy 68 GLVNTADASQSIYQIVDEKSKMLKD-----HGAEVTPNQOITFKALNTYSGEKKI 120
Db 10450 DRQILTTDAAGKPIYSVFNEDGILQPTDSGYAIGHDELVPTESTNGVPLN-----KDG 10504
Oy 121 SPGIYNDQVMVGYVYN 135
Db 10505 TPLPTNDS---GHFV 10516

RESULT 4
AF2097
hypotheoretical protein a112333 (imported) - Nostoc sp. (strain PCC 7120)
C:/Species: Nostoc sp. PCC 7120
A:/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:/Accession: AF2097
R:/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Saemoto, S.; Matsumbe, A.; Itiguchi,
Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takahawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:/Reference number: AB1807; MUID:21595285; PMID:11759840
A:/Accession: AF2097
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-595 <KUR>
A:/Cross-references: UNIPROT:Q8YU11; UNIPARC:UPI00000CE3D3; GB:BA000019; PIDN:BA074032.1;
C:/Experimental source: strain PCC 7120
C:/Genetics:
A:/Gene: a112333

```

```

Query Match 11.6%; Score 81; DB 2; Length 595;
Best Local Similarity 22.9%; Pred. No. 11;
Matches 41; Conservative 28; Mismatches 54; Indels 56; Gaps 11;

Oy 6 TKNFPVST-----TISKSFAPB-PRIOPS-----FGENVKEGALLFSVNLTVPE 50
Db 184 TARIYVOSPTWTVVMQDFGEVAPKIFPAPRYDAALFAGAMENYGAQFLNYG---RLFG 240
Oy 51 NVSQVTVYPVYDEY--GLGRVNTADA-----SQSIYQIVDEKSKMLKDHG 97
Db 241 NLFMIN-WPICGNDYGGVGRILIESAAKGFDFQESRMSQNFHFIQTOLGR-----YG 295
Oy 98 --AEVTPNQOITFKALNY-----TSGEKKISP-----GIYNDQVMVGYVYN 136
Db 296 LAEQVFPHASAFALHPYRESRRLVGLTVREQDILFPGQVASLFPDAIAGNYAN 354

RESULT 5
A81293
probable flagellar hook-associated protein Cj1466 (imported) - Campylobacter jejuni (str
C:/Species: Campylobacter jejuni
C:/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:/Accession: A81293
R:/Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:/Reference number: A81250; MUID:20150912; PMID:10688204
A:/Accession: A81293
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-608 <PAR>
A:/Cross-references: UNIPROT:Q9PMJ4; UNIPARC:UPI00000C1F2D; GB:AL139078; GB:AL111168; NID
A:/Experimental source: serotype O2, strain NCTC 11168
C:/Genetics:
A:/Gene: fliG; Cj1466

Query Match 11.2%; Score 78; DB 2; Length 608;
Best Local Similarity 25.4%; Pred. No. 22;
Matches 34; Conservative 21; Mismatches 39; Indels 40; Gaps 8;

Oy 11 VSTTISKSFPAPBPRF-----QPSFGENVKEGALLFSVNLTVPENVSQVTVYPV---YD 62
Db 215 VSAVASKNEINQDRLDTTIDPBGHYNLSIEG---FSI-----VDGINFHLKLDYD 264
Oy 63 EDYGLRVNTADASQSIYQIVDEKSKMLKDGAETVPO---QITFKALNTYSGEKK 119
Db 265 D-----KNSYSIYVETPDEK---VRDLTAKISGQQLGAQLDLRGRNYSKSE-- 308
Oy 120 ISPGIYNDQVMVGYVYN 133
Db 309 ---GKYEDGIIGGY 319

RESULT 6
JG6010
formate C-acetyltransferase (EC 2.3.1.54) - Clostridium pasteurianum
N:/Alternate names: Pfl protein; pyruvate formate-lyase
C:/Species: Clostridium pasteurianum
C:/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:/Accession: JG6010; PC6004
R:/Weidner, G.; Sauer, G.
J. Bacteriol. 178, 2440-2444, 1996
A:/Title: Molecular characterization of the genes encoding pyruvate formate-lyase and its
A:/Reference number: JG6010; MUID:86218720; PMID:8636053
A:/Accession: JG6010
A:/Molecule type: DNA
A:/Residues: 1-740 <WE11>
A:/Cross-references: UNIPROT:Q46266; UNIPARC:UPI00000131758; EMBL:X93463; NID:g1072360; PI
A:/Accession: PC6004
A:/Molecule type: protein
A:/Residues: 1-740 <WE12>
A:/Cross-references: UNIPARC:UPI0000131758

```

C:Comment: This enzyme must be activated by the lyase-specific pyruvate formate-lyase-activating enzyme (Pfl) and causes peptide cleavage.

C:Gene: pfl

A:Start codon: TTG

C:Complex: homodimer

A:Description: catalyzes the reversible conversion of acetyl-CoA and formate into pyruvate and formate

C:Superfamily: formate C-acetyltransferase 1; glycyl radical homology

A:Pathway: anaerobic glucose metabolism

C:Keywords: acyltransferase; coenzyme A; homodimer; lyase; thiol-ester bond

F/681-740/Domain: glycyl radical homology <GRR>

F/405/Active site: Cys (cysteine thiol radical intermediate) #status predicted

F/406/Active site: Cys (S-acetylcysteine intermediate) #status predicted

F/715/Active site: Gly (stable glycyl radical) #status predicted

Query Match 11.1%; Score 77.5; DB 1; Length 740;

Best Local Similarity 25.0%; Pred. No. 31;

Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

4 IATKN-FPVSTTICKSPFAPRIPQSPFGENVKGGALLFVNLTPENVQV-----TVY 58

336 LVTKNSRRLHTLNLGSAPEPMTVMSENL-PESFKKFAEMSIILDSIQENDDIR 394

59 PVYDEYGL-----GRLVNTADASQSIIVGVDEKGGKMLKHGAETP 102

395 PIVDDVAIACVSAVRGKDMQFGARCNLAKILLINGVDK-----KGIVKP 447

103 N-QOITFKALNTSGEKKISPGIYND-QVMGVYV 136

448 DIEPITDEVLIDY-----EKVKENYFVLEVMAGLYV 479

RESULT 7

Formate acetyltransferase [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C:Accession: E89785

R:Kuroda, M.; Ohno, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mutant, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.

C:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; PMID:21311952; PMID:11418146

A:Accession: E89785

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-749 <KUR>

A:Cross-references: UNIPROT:Q99WZ7; UNIPARC:UPI00000D789B; GB:BA000018; PID:g13700141; F

A:Experimental source: strain N315

C:Gene: pflB

C:Superfamily: formate C-acetyltransferase 1; glycyl radical homology

Query Match 11.1%; Score 77.5; DB 2; Length 749;

Best Local Similarity 23.8%; Pred. No. 31;

Matches 40; Conservative 24; Mismatches 45; Indels 59; Gaps 9;

4 IATKN-FPVSTTICKSPFAPRIPQSPFGENVKGGALLFVNLTPENV-----TVY 52

344 LVTKNSRRLHTLNLGSAPEPMTVMSENL-PESFKKFAEMSIILDSIQENDDIR 394

53 -----SQVTVVPVYDEYGLGRLVNTADASQSIIVGVDEKGGKMLKHGAETP 92

390 TSSIQYENDDIMRESYGGDYGIACCVSAMTTGKMQFGARCNLAKILLINGVDK- 448

93 LKHGAETP-N-QOITFKALNTSGEKKISPGIYNDQVM-----VGVYV 136

DB 449 ---SGAUGVPEGINSEVLEVDYFKKF-----DQMDMLAGYV 467

RESULT 8

AH2515

hypothetical protein alr7304 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C:Accession: AH2515

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iritani, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; PMID:21595285; PMID:11759840

A:Accession: AH2515

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4936 <KUR>

A:Cross-references: UNIPROT:Q8YKJ3; UNIPARC:UPI000011021D; GB:BA000020; PIDN:BA078388.1;

A:Experimental source: strain PCC 7120

C:Gene: alr7304

A:Genome: plasmid

Query Match 11.1%; Score 77.5; DB 2; Length 4936;

Best Local Similarity 25.0%; Pred. No. 3,4e+02;

Matches 42; Conservative 19; Mismatches 48; Indels 59; Gaps 8;

2 TEATKNFPVSTT-----ISKFPAPRIPQSPFGENVKGGALLFVNLTP 47

2406 TGLIKDDPTTVDEGIGSAQWTFPTNLAPERATQYISG-----GLSYKENGITVT 2459

48 VPENVSQVTVTP-----VYDEYGLGRLVNTADASQSIIVGVDEKGGK 91

2460 VPLLSTPIITVYPOAEIYDFHORDVADDPFTNDILET-----SVPSLAIVLRNKGK 2514

92 MKKHGAETPNQOITFKALNTSGEKKI---SPGIYNDQVMGVYV 136

2515 EAKN-----LKITSQPKYDNEKGLIDPQIISFVN 2547

RESULT 9

hypothetical protein T27D1.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T25354

R:Sims, M.

A:Reference number: submitted to the EMBL Data Library, February 1995

A:Accession: T25354

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-356 <WIL>

A:Cross-references: UNIPROT:Q09638; UNIPARC:UPI000013BF59; EMBL:Z48245; PIDN:CAA88290.1;

A:Experimental source: clone T27D1

C:Gene: CESP.T27D1.3

A:Map position: 3

A:Introns: 86/2; 118/2; 164/1; 233/1; 280/2; 322/2

C:Superfamily: vertebrate rhodopsin

Query Match 11.0%; Score 77; DB 2; Length 356;

Best Local Similarity 28.8%; Pred. No. 14;

Matches 19; Conservative 15; Mismatches 30; Indels 2; Gaps 1;

52 VQVTVVPVYDEYGLGRLVNTADASQSIIVGVDEKGGKMLKHGAETPNQOITFK 109

282 VLVNTIHPDIFILMALMTGASNPILYTFESQFRRLRVLCPSDVENETKTY 341

110 ALNTS 115

DB 342 SINNTS 347

RESULT 10

T43531
 probable potassium channel chain n2p2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 09-Jul-2004
 C:Accession: T43531; T22711
 R:Wang, Z.W.; Salkoff, L.
 submitted to the EMBL Data Library, August 1998
 A:Description: Potassium channels in C. elegans.
 A:Reference number: 222450
 A:Accession: T43531
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-569 <MAN>
 A:Cross-references: UNIPROT:Q20817; UNIPARC:UPI000007FPd6; EMBL:AF083654; PIDs:AAC32865-
 R:Harria, B.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: 219604
 A:Accession: T22711
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-569 <ML>
 A:Cross-references: UNIPARC:UPI000007FPd6; EMBL:Z78198; PIDs:CA801568.1; GSPDB:GN00023;
 A:Experimental source: clone F55C5
 A:Gene: F55C5.3
 A:Map position: 5
 A:Introns: 43/3; 140/1; 171/2; 294/2; 338/3; 382/2; 508/2

Query Match 10.9%; Score 76.5; DB 2; Length 569;
 Best Local Similarity 30.8%; Pred. No. 27;
 Matches 20; Conservative 12; Mismatches 28; Indels 5; Gaps 1;

DB 56 TTYPYDEDEYGLRVNTADASQSIITYOYDEKGMKLDHGAETVPCNOITFKALNTYS 115
 468 TMSIYDEDEEGVKRLIQSOSLEDRIIFLFDHKKSMLEKRTTYPS-----KADKYQ 522

QY 116 GEKXI 120
 DB 523 TQGRV 527

RESULT 11
 F69481
 probable acid-CoA ligase (EC 6.2.1.-) - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #ext_change 09-Jul-2004
 C:Accession: F69481
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Klinkner, E.F.;
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: F69481
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-557 <KLF>
 A:Cross-references: UNIPROT:Q28423; UNIPARC:UPI0000056B85; GB:AE000975; GB:AE000782; NID
 C:Superfamily: probable acyl-CoA ligase medium chain; acetate-CoA ligase homology
 C:Keywords: acid-thiol ligase
 F:78-546/Domain: acetate-CoA ligase homology <ACL>

Query Match 10.9%; Score 76; DB 2; Length 557;
 Best Local Similarity 27.2%; Pred. No. 30;
 Matches 31; Conservative 22; Mismatches 31; Indels 30; Gaps 7;

DB 4 IATKAPVSTTIS--KSFFA-----PDP-----RIPS--FGENVKRGALLFSV 44
 295 VALKHFMASQSLFDLTKTFAQGSAPPPWTFTVFLKNRGEIENINWGQ---EGTGLFSY 351

QY 45 NLTVENVSQVTVYVDEYDGLRVNTADASQSIITYOYDEKGMKLDHGA 98

DB 352 DRTIPDLKRAKSPPIPK-----VADLPFPAIEIKIYDESGKEM-KEPGS 397

RESULT 12
 T99882
 pectin methyl-esterase-like protein - Arabidopsis thaliana
 N:Alternate names: protein T211.140
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #ext_change 09-Jul-2004
 A:Accession: T99882
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 224493
 A:Accession: T99882
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-361 <BEV>
 A:Cross-references: UNIPROT:Q9LY17; UNIPARC:UPI0000048AB; EMBL:AL163912; GSPDB:GN00063;
 A:Experimental source: cultivar Columbia; BAC clone T211
 C:Genetics:
 A:Gene: ATSP:T211.140
 A:Map position: 5
 A:Introns: 68/3; 161/3; 229/3; 309/2

Query Match 10.8%; Score 75.5; DB 2; Length 361;
 Best Local Similarity 28.0%; Pred. No. 19;
 Matches 28; Conservative 16; Mismatches 23; Indels 33; Gaps 6;

QY 48 VPENVSV-----TVYVYDEYDGLRVNTADASQSIITYOYDEKGMKLDHGAETV 101
 DB 28 IPENRAQIPQWFKTNVKSQRKGLTLPALRAEAARQII--TVNQKG-----GA--- 75

QY 102 PNOQITFKALN-----YTSGEK-----KISPGIYNDQYV 131
 DB 76 -----NFKTLBAIKSTFTGKNKRVYITLACGVNNEKTYI 110

RESULT 13
 A82851
 chaperone protein precursor XF0082 (imported) - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 09-Jul-2004
 C:Accession: A82851
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: For a complete list of authors see reference number A59328 below
 A:Accession: A82851
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-266 <SIM>
 A:Cross-references: UNIPROT:Q9PH64; UNIPARC:UPI00000C22C4; GB:AE003862; GB:AE003849; NID
 A:Experimental source: strain 9a5c
 R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.; A
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carreiro, D.M.; Carreir, E
 de-Neto, E.; Docena, C.; El-Dorrry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohn
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajim, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizy
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; de Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tanaka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 A:Genetics:
 A:Gene: XF0082
 C:Superfamily: chaperone protein papp

Query Match 10.7%; Score 75; DB 2; Length 266;
Best Local Similarity 22.9%; Pred. No. 15;
Matches 41; Conservative 22; Mismatches 44; Indels 72; Gaps 9;
QY 16 SXSFPAPF---RIGPSFGENGKEGALLFSVNLTPF-----NVSQVIVPYDD 64
DB 78 SAAFLVTPPIRVEPVGSQL---RLFFGTITLPEDESLFVNLVLEIPPTLEHSD 133
QY 65 -----YGLGRVLTADASQSI 81
DB 134 KNAOTSADKSDSQNLQALFRSRIRKVFYRPALKGTANSPASLSKVNQSLVSNPT 193
QY 82 -YQI-----VDEKGMKLKDH-GAEVTPNQITPKALNTSG---EKISPGIYND 127
DB 194 PYHVLTRTEALDDKGRKETIDKIGVLLTPGQEVTFKA-NNDGLSSRKVSFSFYIND 251
RESULT 14
D64534
DNA topoisomerase I - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #ext_change 09-Jul-2004
C:Accession: D64534
R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodex, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Wetman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:9794467; PMID:9252185
A:Accession: D64534
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-736 <TOM>
A:Cross-references: UNIPROT:P55991; UNIPARC:UPI000013713A; GB:AE000533; GB:AE000511; NID
C:Superfamily: DNA topoisomerase I
Query Match 10.7%; Score 75; DB 2; Length 736;
Best Local Similarity 19.7%; Pred. No. 53;
Matches 23; Conservative 29; Mismatches 63; Indels 2; Gaps 1;
QY 1 RTEITKMPVSTTSKSFAPFRIQPSFGENGKEGALLFSVNLTPENVSQVTPV 60
DB 228 KNELEKESYAISSIVKSKSPPT--PPMTSTLQOSASLLGFSPKTMSIAOKLYEGV 285
QY 61 YDEDEGLRLVLTADASQSIYQIVDEKGMKLKDHGAETPNQITPKALNTSGE 117
DB 286 ATPGCMGVITTYMRTDSLNIKALEARNKILKDYKDYLPKAKVYSSKNKMAOE 342
RESULT 15
H82802
Embrial assembly protein XF0478 (imported) - Xylella fastidiosa (strain 945C)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 09-Jul-2004
C:Accession: H82802
R:Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: H82802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1472 <SIM>
A:Cross-references: UNIPROT:Q9PC24; UNIPARC:UPI00000C2413; GB:AE003897; GB:AE003849; NID
A:Experimental source: strain 945C
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Ribeiro, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; From
J.D.; Junqueira, M.L.; Kemper, B.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0478
Query Match 10.7%; Score 75; DB 2; Length 1472;
Best Local Similarity 22.7%; Pred. No. 1.3e+02;
Matches 34; Conservative 22; Mismatches 40; Indels 54; Gaps 7;
QY 11 VSTTSKSFAPFRIQPSF-----GENV-----GKEGALLFSVNLTPENVS 53
DB 790 VGASVSTSF-----VIRISSTHDGKMTGEMTYRIRIDANGMLGDVLSASAGVPSGTS 845
QY 54 QTVYVYDEDEYGLRLVLTADASQSIYQIVDEK-----GK 90
DB 846 ALAKRVIT---ALSHVDITRA--SVRRFVAKELVSSSGDVATDAQVGRFGYTPR 900
QY 91 KMLKDHGAETPNQITPKALNTSGEKKI 120
DB 901 GVIDDFGSSVTPNOLV-----NYLRGDKSM 925

Search completed: April 10, 2006, 10:21:26
Job time : 40 secs

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```

1 STREET: 100 Beaver Street
2 CITY: Waltham
3 STATE: Massachusetts
4 COUNTRY: USA
5 ZIP: 02354
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: CD-ROM ISO9660
9 COMPUTER: PC
10 OPERATING SYSTEM: <Unknown>
11 SOFTWARE: ASCII
12
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/09/107,532A
15 FILING DATE: 30-Jun-1998
16
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: 60/085,598
19 FILING DATE: 14 May 1998
20 APPLICATION NUMBER: 60/051571
21 FILING DATE: July 2, 1997
22
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Artidello, Pamela; Deneka
25 REGISTRATION NUMBER: 40,489
26 REFERENCE/DOCKET NUMBER: GTC-012
27
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (781)893-5007
30 TELEFAX: (781)893-8277
31
32 INFORMATION FOR SEQ ID NO: 4742:
33
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 378 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39 HYPOTHETICAL: YES
40 ORIGINAL SOURCE:
41 ORGANISM: Enterococcus faecium
42
43 FEATURE:
44 NAME/KEY: misc feature
45 LOCATION: (B) LOCATION 1...378
46 SEQUENCE DESCRIPTION: SEQ ID NO: 4742:
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
97
```

```

Query Match 11.2% Score 78.5; DB 2; Length 378;
Best Local Similarity 27.1%; Pred No. 0.56;
Matches 35; Conservative 19; Mismatches 48; Indels 27; Gaps 7

QY 13 TTISKFFAPBPRIQPSFGENVGK-EGALLPSVLTVPBNVSQVTVY-----PYY 61
DB 176 TSIGHGRTPBAGAQPAIGK-VGKTEQVQEAKEVILPETIEKQVYQAMRSAPYECPAY 234
QY 62 D-----EYIGGLRLVNTDASGSIYYQIVDEGKKMLKHGAEVY--PQQITFK-- 109
DB 235 DLFAIDEPVEMFGLGRV--GELPEITIEAFVEQVKEAFOLDGLRIQVQKAKSSVKRI 291
QY 110 ALNTYSGEK 118
DB 292 AICGGSGEK 300

RESULT 3
US-09-323-872A-25
; Sequence 25, Application US/09323872A
; Patent No. 6395539
; GENERAL INFORMATION:
; APPLICANT: Coochigano, Peter
; TITLE OR INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: OHU-03640
; CURRENT APPLICATION NUMBER: US/09/323, 872A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/072,433
; PRIOR FILING DATE: 1998-05-04
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
LENGTH: 740

```

```

; TYPE: PRT
; ORGANISM: Clostridium pasteurianum
US-09-323-872A-25

Query Match      11.1%; Score 77.5; DB 2; Length 740;
Match Local Similarity 25.0%; Pred. No. 2.1;
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8

Qy   4 IATKN-PVVSSTTICKSFAPRPIQPSFGSEVGKCALPFSVLTVDENNSQV-----TVY 58
      : ||| :                : ||| :          : :: :
Db   336 LVTGNSRYLHTLTLLNGASAPENMTVMSENL -PESFRKKFAEMSLITDSIQENDDLMR 394
      : ||| :                : ||| :          : :: :

Qy   59 PVYDEDYGL-----GRLVNTADASQSITIYIVDEKGGKMLKDHCAGETVP 102
      : : | :                : : | :          : ||| :
Db   395 PIYDDPDVAICCVSAMREVGMQDFPGARCNIACILLLAINGVDEK-----KGIRKVP 447
      : : | :                : : | :          : ||| :

Qy   103 N-QOIFKALNYTSGEKKISPGIYN-DQWVVGYVYN 136
      : : | :                : : | :          : ||| :
Db   448 DIEPTDVALDY---EKVKENYKYULEYMAGLYYN 479
      : : | :                : : | :          : ||| :

```

[illegible]

QY 52 VSGVTVPVYDDE-----YGLRLVNTADASGSI IQYLVDKSKMLKHGAETPNO 104
| : | : | : ||| : : : |
DB 253 VMNGYLPRKYLLKDTFPNDGVFQTGDIAIDDEGYIITY---DRRKDLIIISGENIYP-Y 307

D_b

QY 105 QITFKALNTTSEKKISPGIYN---QVWNGYYV 135
|| : | : || : |||
308 QIETIAKDFEGIEDAVCGVISDDPMGWGVPIILYYV 341

RESULT 9
US-09-134-001C-5114
; Sequence 5114, Application US/09134001C
; Patent No. 6380170
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 5114
 LENGTH: 497

Query Match Similarity 10.7% Score 74.5; DB 2; Length 497;
 Best Local Similarity 24.7% Pred No. 2.7;
 Matches 38; Conservative 21; Mismatches 58; Indels 37; Gaps 7

OY 8 NFPVSTTISKFFPAPEPRIQPSGENVGK-----EGALLFSLNLTIPEN 51
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 296 SFGMTETCSQ-FLTASPQLMKERFDTVGKPSENVVKIKNPNAVGHGELLIK----GEN 34

```

qy      52 VSGVTVYPIYDED-----XGLGRLVNTADASQSIITYIVDEKGMKLKDHAAEVTNQ 104
       | :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     350 VMNGVLYPKYLKDFDNDGYFQTGDLAIEDBGVIYY---DRKKLIISGENIYP-Y 404

```

Qy 105 QITPKALNTSGEKKISPGIYND---QVMGVYV 135
 |||::|||:|||
 Db 405 QIETIAKDPEGIEDAVCGVISDDTWGQVPIIYV 438

RESULT 10
US-09-248-796A-17536

GENERAL INFORMATION:
APPLICANT: Ketch Welnstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 17536

Query Match	10.6%	Score	74	DB	2	Length	467
Best Local Similarity	25.7%	Pred. NO.	2.8				
Matches	28	Conservative	22	Mismatches	37	Indels	22
						Gaps	4

QY 5 ATKNFVSTTISKSPFAPEPRQPSFGENGKEGALLFSVNLTPENNQSQVTPYVDED 64

Db 252 ASGNIPVAGTAKKAPSTDENLAHISF-----BAGILENPTTTPPKMMKLYDP----- 3000

Qy 65 YGLGRVMTADASQSIITQYIDKGS---KMYLKDHGAELVTPNQOITFKA 110

Db 301 -----TDAPDKPET--FSVVPFEKFLPKYLLIDDSKKEITPVPALLFEA 341

RESULT 11
US-09-328-352-6886
; Sequence 6886, Application US/09328352
; Patent No. 6562958

```

1  APPLICANT: Gary L. Breton et al.
2  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
3  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
4  FILE REFERENCE: GTC099-03PA
5  CURRENT APPLICATION NUMBER: US/09/328,352
6  CURRENT FILING DATE: 1999-06-04
7  NUMBER OF SEQ ID NOS: 8252
8  SEQ ID NO 6886
9  LENGTH: 218
10 TYPE: PRT
11 ORGANISM: Acinetobacter baumannii
12 JS-09-328-352-6886

```

Query Match	10.3%	Score 72;	DB 2;	Length 218;
Best Local Similarity	25.0%	Pred. No. 1,6;		
Matches 36;	Conservative 20;	Mismatches 68;	Indels 20;	Gaps 6

Dy 4 IATKNFPSTTISKSFAPBPRQPSFGENVK-EGALLSEVNLTPEINISQVTV--PV 60
 : | | : : | | : | :
Db 70 LVTNTCTIDTGKNLTVNLPTVSSQLSKNAGDVAGRPFQINLTCNCASGVKVAITYEPC 129

61 YDEYDGLRLVN---TADASQSIITYOI-----VDEGKKMLKDHAETV-----PNOQ 105
 130 ATVDFTNGRLNQATSGAANVTIQLGSNNAVIPVIAGSGAGADNSQWTVNVSAGGNAD 189

```

OY      106  ITFKALNTSGEKKISPGIYNDQV 123
          : : | : |
Db      190  LNYAEYATGAS--TAGVTSQV 211

```

RESULT 12
US-09-949-016-6079
; Sequence 6079, Application US/09949016

```

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C0001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR APPLICATION DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: PASTESEQ for Windows Version 4.0
SEQ ID NO 6079
LENGTH: 1005
TYPE: PRT
ORGANISM: Human
IS-09-949-016-6079

```

Query Match	10.3%;	Score 72;	DB 2;	Length 1005;
Best Local Similarity	37.5%;	Pred. No. 16;		
Matches 21;	Conservative	8;	Mismatches 23;	Indels 4;
				Gaps 2;

by I R E I A T K N P E V S T I I S K S F F A P E P L O P S F G E N V G K G A L L F S V N L I V E N S Q V I 56

Db 886 RTEISTKEVPVQTETKTITTESPQIDGAG---GDSGLTLTAQIT-SESVSTT 937

RESULT 13

US-09-949-016-7359
; Sequence 7359, Application US/09949016
; Patent No. 6812319
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO010307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7359
; LENGTH: 1008
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7359

Query Match 10.3%; Score 72; DB 2; Length 1008;

Best Local Similarity 37.5%; Pred. No. 16; Mismatches 23; Indels 4; Gaps 2;
Matches 21; Conservative 8;

Db 889 RTEISTKEVPVQTETKTITTESPQIDGAG---GDSGLTLTAQIT-SESVSTT 940

RESULT 14

US-09-134-000C-6740
; Sequence 6740, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucet et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6740
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6740

Query Match 10.2%; Score 71.5; DB 2; Length 1467;
Best Local Similarity 25.2%; Pred. No. 33; Mismatches 36; Indels 41; Gaps 7;
Matches 33; Conservative 21;

Db 835 KIEBSFTEINKE--VEKTIIVVEEQIEKVEKKKTTEDVDRLHNGFARTIAPFLMA 892

Qy 65 YGL--GRI---VNTADASQSIITVOIVDEKSKMLKDHGAEVTPNQITTFKALNTSGEK 118

Db 893 YGVEDTRLENFENVIDEATFEDLTSTITTEFKR-LRD-GFEYPDN-----NGET 939

Qy 119 KISPGIYNDQV 129

Db 940 KVIPLGFNEVV 950

RESULT 15
US-09-538-092-213

; Sequence 213, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Tracy A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurataseqFormatter Version 0.9
; SEQ ID NO 213
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YDR480W
US-09-538-092-213

Query Match 10.1%; Score 70.5; DB 2; Length 323;

Best Local Similarity 24.7%; Pred. No. 4.4; Mismatches 26; Indels 23; Gaps 5;
Matches 23; Conservative 21;

Qy 49 PENVSQ---VTVPVYDEYDYGRLVNTADASQSIITVOIVDEKSKML-----KDHG 97

Db 22 PRNQQLMLVSGDLDEDRLEKMKISIKKSRDI-----EKQKLLISRLSQKEEDHS 75

Qy 98 AE---VTNQO--ITFKALNTSGEKKISPGI 124

Db 76 GKPPITTSFAEKTVPFKSLNHSIKRKRVPPAL 108

Search completed: April 10, 2006, 10:22:19
Job time : 47 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:21:40 ; Search time 166 Seconds
(without alignments)
342.318 Million cell updates/sec

Title: US-10-768-093-9
Perfect score: 699
Sequence: 1 RTEIATKPNFVSTTISKSPF.....EKKISPGIYNDQVWGYVN 136

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	699	100.0	136 4 US-10-768-093-9	Sequence 9, Appli
2	699	100.0	136 5 US-10-754-641-9	Sequence 9, Appli
3	699	100.0	154 4 US-10-768-093-5	Sequence 5, Appli
4	699	100.0	154 5 US-10-754-641-5	Sequence 5, Appli
5	113.5	16.2	146 4 US-10-768-093-10	Sequence 10, Appli
6	113.5	16.2	167 4 US-10-768-093-6	Sequence 6, Appli
7	113.5	16.2	167 5 US-10-754-641-6	Sequence 25, Appli
8	77.5	11.1	740 4 US-10-267-989-25	Sequence 25, Appli
9	77.5	11.1	740 4 US-10-357-567-25	Sequence 25, Appli
10	76	10.9	557 4 US-10-369-493-21384	Sequence 21384, A
11	76	10.9	2358 4 US-10-282-122A-68607	Sequence 68607, A
12	75.5	10.8	246 4 US-10-282-122A-67315	Sequence 67315, A
13	75	10.7	640 4 US-10-282-122A-44691	Sequence 44691, A
14	75	10.7	736 3 US-09-815-242-11293	Sequence 11293, A
15	75	10.7	736 4 US-10-282-122A-58624	Sequence 58624, A
16	74.5	10.7	249 4 US-10-424-599-250712	Sequence 250712, A
17	74.5	10.7	497 4 US-10-724-972A-6030	Sequence 6030, Ap
18	74.5	10.7	750 5 US-10-501-282-3988	Sequence 3988, Ap
19	73.5	10.5	157 4 US-10-424-599-190153	Sequence 190153, A
20	73.5	10.5	357 4 US-10-282-122A-66978	Sequence 66978, A
21	73.5	10.5	356 4 US-10-282-122A-78182	Sequence 78182, A
22	73	10.4	302 4 US-10-335-977-7475	Sequence 7475, Ap
23	73	10.4	541 4 US-10-335-977-7476	Sequence 7476, Ap
24	72.5	10.4	736 3 US-09-815-242-11473	Sequence 11473, A
25	72.5	10.4	645 4 US-10-425-114-71452	Sequence 71452, A
26	72.5	10.4	1773 6 US-11-097-143-18516	Sequence 18516, A
27	72	10.3	274 4 US-10-365-493-8865	Sequence 8865, Ap

28	72	10.3	398 3 US-09-925-301-1264	Sequence 1264, Ap
29	72	10.3	398 4 US-10-264-049-2703	Sequence 2703, Ap
30	72	10.3	505 4 US-10-437-963-110397	Sequence 110397, A
31	72	10.3	520 4 US-10-369-493-11092	Sequence 11092, A
32	72	10.3	1005 5 US-10-921-707-1	Sequence 1, Appli
33	72	10.3	1705 4 US-10-437-963-118350	Sequence 118350, A
34	71.5	10.2	357 5 US-10-732-923-10066	Sequence 10066, A
35	71.5	10.2	357 5 US-10-732-923-10067	Sequence 10067, A
36	71.5	10.2	425 4 US-10-424-599-168174	Sequence 168174, A
37	71.5	10.2	571 4 US-10-282-122A-53248	Sequence 53248, A
38	71.5	10.2	657 4 US-10-282-122A-45312	Sequence 45312, A
39	71.5	10.2	1207 4 US-10-437-963-111041	Sequence 111041, A
40	70.5	10.1	357 3 US-09-815-242-10981	Sequence 10981, A
41	70.5	10.1	357 4 US-10-260-877-128	Sequence 128, App
42	70.5	10.1	357 4 US-10-268-611-41	Sequence 41, Appli
43	70.5	10.1	357 4 US-10-282-122A-58099	Sequence 58099, A
44	70.5	10.1	357 5 US-10-824-194-50	Sequence 50, Appli
45	70.5	10.1	430 4 US-10-282-122A-60954	Sequence 60954, A

ALIGNMENTS

RESULT 1
US-10-768-093-9
Sequence 9, Application US/10768093
Publication No. US20040156829A1
GENERAL INFORMATION:
APPLICANT: Wolf, Marcia K
Cassels, Fred J
Boedeker, Edgar C
TITLE OF INVENTION: Transformed Bacteria Producing Gs6 Antigens as Vaccines
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hendricks and Assoc
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/768, 093
FILING DATE: 02-Feb-2004
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/479, 877B
FILING DATE: 10-Jan-2000
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna M
REGISTRATION NUMBER: 32,535
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/425-8405
TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-768-093-9
Query Match 100.0%; Score 699; DB 4; Length 136;
Best Local Similarity 100.0%; Pred. No. 5e-67;

6/6/05

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

DB 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

QY 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120
61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

DB 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

QY 121 SPGIYNDQVMGYVYN 136
121 SPGIYNDQVMGYVYN 136

DB 121 SPGIYNDQVMGYVYN 136

RESULT 2
US-10-754-641-9
Sequence 9, Application US/10754641
Publication No. US20050025787A1
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research and Materiel Command
APPLICANT: Wolf, Marcia K.
APPLICANT: Cassels, Frederick J.
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES
FILE REFERENCE: 034047.033.3
CURRENT APPLICATION NUMBER: US/10/754,641
CURRENT FILING DATE: 2004-01-12
PRIOR APPLICATION NUMBER: 09/479,877
PRIOR FILING DATE: 2000-01-10
PRIOR APPLICATION NUMBER: 08/788,145
PRIOR FILING DATE: 1997-01-24
PRIOR APPLICATION NUMBER: 08/243,482
PRIOR FILING DATE: 1994-05-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.2
SEQ ID NO: 9
LENGTH: 136
TYPE: PR
ORGANISM: Escherichia coli
US-10-754-641-9

Handwritten signature/initials.

Query Match 100.0%; Score 699; DB 5; Length 136;
Best Local Similarity 100.0%; Pred. No. 5e-67;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

DB 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

QY 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120
61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

DB 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

QY 121 SPGIYNDQVMGYVYN 136
121 SPGIYNDQVMGYVYN 136

DB 121 SPGIYNDQVMGYVYN 136

RESULT 3
US-10-768-093-5
Sequence 5, Application US/10768093
Publication No. US20040156829A1
GENERAL INFORMATION:
APPLICANT: Wolf, Marcia K.
APPLICANT: Cassels, Fred J.
APPLICANT: Boedeker, Edgar C.
TITLE OF INVENTION: Transformed Bacteria Producing GS6
Antigens as Vaccines
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hendricks and Assoc
STREET: P.O. Box 2509

CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/768,093
FILING DATE: 02-Feb-2004
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/479,877B
FILING DATE: 10-Jan-2000
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenn M.
REGISTRATION NUMBER: 32,535
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/425-8405
TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 154 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: unknown
TOPOLOGY: unknown
HYPOTHEICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-768-093-5

Query Match 100.0%; Score 699; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 6e-67;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

DB 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENVSQVTVPV 60

QY 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120
61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

DB 61 YDEDYGLRLVNTADASQSIYIYIVDEKGMKLDHGAETVPNOQITFPALNTSGEKKI 120

QY 121 SPGIYNDQVMGYVYN 136
121 SPGIYNDQVMGYVYN 136

DB 121 SPGIYNDQVMGYVYN 154

RESULT 4
US-10-754-641-5
Sequence 5, Application US/10754641
Publication No. US20050025787A1
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research and Materiel Command
APPLICANT: Wolf, Marcia K.
APPLICANT: Cassels, Frederick J.
APPLICANT: Boedeker, Edgar C.
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES
FILE REFERENCE: 034047.033.3
CURRENT APPLICATION NUMBER: US/10/754,641
CURRENT FILING DATE: 2004-01-12
PRIOR APPLICATION NUMBER: 09/479,877
PRIOR FILING DATE: 2000-01-10
PRIOR APPLICATION NUMBER: 08/788,145
PRIOR FILING DATE: 1997-01-24
PRIOR APPLICATION NUMBER: 08/243,482
PRIOR FILING DATE: 1994-05-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.2
SEQ ID NO: 5

LENGTH: 154
TYPE: PRT
ORGANISM: Escherichia coli
US-10-754-641-5

Query Match 100.0%; Score 699; DB 5; Length 154;
Best Local Similarity 100.0%; Pred. No. 6e-67;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKPFVSTTSSKSPFAPEPIOPSPFGENVGKGLLPSVNLTVPENVSQVTVYP 60
DB 19 RTEIATKPFVSTTSSKSPFAPEPIOPSPFGENVGKGLLPSVNLTVPENVSQVTVYP 78
QY 61 YDEIDYGLRLVNTADASQSIYQIVDEKGGKMLKDHGAEVTPNQITFKALNTYSGEKKI 120
DB 79 YDEIDYGLRLVNTADASQSIYQIVDEKGGKMLKDHGAEVTPNQITFKALNTYSGEKKI 138
QY 121 SPGIYNDQVMVGYVYN 136
DB 139 SPGIYNDQVMVGYVYN 154

RESULT 5

US-10-768-093-10
Sequence 10, Application US/10768093
Publication No. US20040156829A1
GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K

Cassels, Fred J

Boedecker, Edgar C

TITLE OF INVENTION: Transformed Bacteria Producing G86

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Hendricks and Assoc

STREET: P.O. Box 2509

CITY: Fairfax

STATE: VA

COUNTRY: US

ZIP: 22031

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/768,093

FILING DATE: 02-Feb-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/479,877B

FILING DATE: 10-Jan-2000

ATTORNEY/AGENT INFORMATION:

NAME: Hendricks, Glenna M

REGISTRATION NUMBER: 32,535

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703/425-8405

TELEFAX: 703/425-8406

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-768-093-10

Query Match 16.2%; Score 113.5; DB 4; Length 146;
Best Local Similarity 28.5%; Pred. No. 0.00061;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

6/11/06

QY 7 KNFPVSTTSSKSPFAPE---PRIOPSPFGENVGKGLLPSVNLTVPENVSQVTVYP--- 59
DB 6 KSLDVNVIQNF-LPDISAVRIIPVNYDSDPKLSQLYTVEMTIPAGVSAVXIAPIDS 64
QY 60 YDEIDYGLRLVNTADASQSIYQIVDEKGGKMLKDHA-----EVPNQITF 108
DB 65 LTSSGQIGKLVNANNPDQNNNYI-----RKDSAGNFMAGQKSPFKENTSYTF 116
QY 109 KALNTYSGE---KKISPGIYNDQVMVGYVYN 136
DB 117 SAI-YTGEIYNSGYSSTGTAGNLTVSFSYN 146

RESULT 6

US-10-768-093-6
Sequence 6, Application US/10768093
Publication No. US20040156829A1
GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K

Cassels, Fred J

Boedecker, Edgar C

TITLE OF INVENTION: Transformed Bacteria Producing G86

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Hendricks and Assoc

STREET: P.O. Box 2509

CITY: Fairfax

STATE: VA

COUNTRY: US

ZIP: 22031

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/768,093

FILING DATE: 02-Feb-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/479,877B

FILING DATE: 10-Jan-2000

ATTORNEY/AGENT INFORMATION:

NAME: Hendricks, Glenna M

REGISTRATION NUMBER: 32,535

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703/425-8405

TELEFAX: 703/425-8406

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 167 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-768-093-6

Query Match 16.2%; Score 113.5; DB 4; Length 167;
Best Local Similarity 28.5%; Pred. No. 0.00074;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

6/11/06

QY 109 KALNTYSGE---KXISPGIYNDQVWVGYYVN 136
DB 138 SAI-YTGGEYRNSGYSSGTYAGNLTVSFYSN 167

RESULT 7
US-10-754-641-6
Sequence 6, Application US/10754641
Publication No. US20050025787A1

GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research and Materiel Command
APPLICANT: Wolf, Marcia K.
APPLICANT: Casadei, Frederick J.
APPLICANT: Boedeker, Edgar C.
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES
FILE REFERENCE: 034047 013.3
CURRENT APPLICATION NUMBER: US/10/754,641
CURRENT FILING DATE: 2004-01-12
PRIOR APPLICATION NUMBER: 09/479,877
PRIOR FILING DATE: 2000-01-10
PRIOR APPLICATION NUMBER: 08/788,145
PRIOR FILING DATE: 1997-01-24
PRIOR APPLICATION NUMBER: 08/243,482
PRIOR FILING DATE: 1994-05-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 167
TYPE: PRT
ORGANISM: Escherichia coli
US-10-754-641-6

Query Match 16.2%; Score 113.5; DB 5; Length 167;
Best Local Similarity 28.5%; Pred. No. 0.00074;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

QY 7 KNFVSTTISKSPFAPE---PRIQSFGENYKEGALLFSVNLTVPENVSQVYTP--- 59
DB 27 KSLDVNVNIEQNFI-PDIDSAVRILIPVNYDSPKLDLSQLYTEMTIPAGVSAVKIAPTDS 85
QY 60 VYDEYDGLVNTADASQSIYQIVDEKGMKLDHGA-----EYTPNOITF 108
DB 86 LTSSQOQIGKLVNPNPDNMNYI-----RKDSGAGNFMAGKSGSPVKNSTYTF 137
QY 109 KALNTYSGE---KXISPGIYNDQVWVGYYVN 136
DB 138 SAI-YTGGEYRNSGYSSGTYAGNLTVSFYSN 167

RESULT 8
US-10-267-989-25
Sequence 25, Application US/10367989
Publication No. US20030199035A1
GENERAL INFORMATION:
APPLICANT: ARNAU, Jose
APPLICANT: VRANG, Astrid
APPLICANT: ISRAELSEN, Hans
APPLICANT: JOERGENSEN, Flemming
APPLICANT: MADSEN, Soeren
TITLE OF INVENTION: METABOLICALLY ENGINEERED LACTIC ACID BACTERIA AND
TITLE OF INVENTION: MEANS FOR PROVIDING
FILE REFERENCE: ARNAU-1A
CURRENT APPLICATION NUMBER: US/10/267,989
CURRENT FILING DATE: 2002-10-10
PRIOR APPLICATION NUMBER: US/08/981,097
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: PCT/DK97/00336
PRIOR FILING DATE: 1997-08-20
PRIOR APPLICATION NUMBER: 08/701,458
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn version 3.0
SEQ ID NO 25
LENGTH: 740
TYPE: PRT
ORGANISM: Clostridium pasteurianum
US-10-267-989-25

Query Match 11.1%; Score 77.5; DB 4; Length 740;
Best Local Similarity 25.0%; Pred. No. 44;
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

QY 4 IATKN-PVSTTISKSPFAPEPRIQSFGENYKEGALLFSVNLTVPENVSQV---TVY 58
DB 336 LVTKNSFRYLHTLNLMSAPBNMTVMSENV-PBSFKKFAEMSLTDSIOYENDDIR 394
QY 59 PUYDEDYGL-----GRLVNTADASQSIYQIVDEKGMKLDHGAETVP 102
DB 395 PIYGDVAILACCVSAMRYGKMQFFGARNCLAKCLLALINGVDK-----KXIKVVP 447
QY 103 N-QQITFPALNTSGEKKISPGIYND-QVWVGYYVN 136
DB 448 DIEPITDEVLDY---EKVKENYFVLEVMAGLYVN 479

RESULT 9
US-10-357-567-25
Sequence 25, Application US/10357567
Publication No. US20040038382A1
GENERAL INFORMATION:
APPLICANT: Coschigano, Peter
TITLE OF INVENTION: Compositions and Methods for Bioremediation
FILE REFERENCE: OHU-07748
CURRENT APPLICATION NUMBER: US/10/357,567
CURRENT FILING DATE: 2003-02-04
PRIOR APPLICATION NUMBER: 09/072,433
PRIOR FILING DATE: 1998-05-04
PRIOR APPLICATION NUMBER: 60/046,845
PRIOR FILING DATE: 1997-05-05
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 25
LENGTH: 740
TYPE: PRT
ORGANISM: Clostridium pasteurianum
US-10-357-567-25

Query Match 11.1%; Score 77.5; DB 4; Length 740;
Best Local Similarity 25.0%; Pred. No. 44;
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

QY 4 IATKN-PVSTTISKSPFAPEPRIQSFGENYKEGALLFSVNLTVPENVSQV---TVY 58
DB 336 LVTKNSFRYLHTLNLMSAPBNMTVMSENV-PBSFKKFAEMSLTDSIOYENDDIR 394
QY 59 PUYDEDYGL-----GRLVNTADASQSIYQIVDEKGMKLDHGAETVP 102
DB 395 PIYGDVAILACCVSAMRYGKMQFFGARNCLAKCLLALINGVDK-----KXIKVVP 447
QY 103 N-QQITFPALNTSGEKKISPGIYND-QVWVGYYVN 136
DB 448 DIEPITDEVLDY---EKVKENYFVLEVMAGLYVN 479

RESULT 10
US-10-369-493-21384
Sequence 21384, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

```

1  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
2  /
3  / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
4  /
5  / FILE REFERENCE: 38-10 (52052)B
6  /
7  / CURRENT APPLICATION NUMBER: US/10/369,493
8  /
9  / CURRENT FILING DATE: 2003-02-28
10 /
11 / PRIOR APPLICATION NUMBER: US 60/360,039
12 /
13 / PRIOR FILING DATE: 2002-02-21
14 /
15 / NUMBER OF SEQ ID NOS: 47374
16 /
17 / SEQ ID NO 21384
18 /
19 / LENGTH: 557
20 /
21 / TYPE: PRT
22 /
23 / ORGANISM: Archaeoglobus fulgidus
24 /
25 /
26 /
27 /
28 /
29 /
30 /
31 /
32 /
33 /
34 /
35 /
36 /
37 /
38 /
39 /
40 /
41 /
42 /
43 /
44 /
45 /
46 /
47 /
48 /
49 /
50 /
51 /
52 /
53 /
54 /
55 /
56 /
57 /
58 /
59 /
60 /
61 /
62 /
63 /
64 /
65 /
66 /
67 /
68 /
69 /
70 /
71 /
72 /
73 /
74 /
75 /
76 /
77 /
78 /
79 /
80 /
81 /
82 /
83 /
84 /
85 /
86 /
87 /
88 /
89 /
90 /
91 /
92 /
93 /
94 /
95 /
96 /
97 /
98 /
99 /
100 /
101 /
102 /
103 /
104 /
105 /
106 /
107 /
108 /
109 /
110 /
111 /
112 /
113 /
114 /
115 /
116 /
117 /
118 /
119 /
120 /
121 /
122 /
123 /
124 /
125 /
126 /
127 /
128 /
129 /
130 /
131 /
132 /
133 /
134 /
135 /
136 /
137 /
138 /
139 /
140 /
141 /
142 /
143 /
144 /
145 /
146 /
147 /
148 /
149 /
150 /
151 /
152 /
153 /
154 /
155 /
156 /
157 /
158 /
159 /
160 /
161 /
162 /
163 /
164 /
165 /
166 /
167 /
168 /
169 /
170 /
171 /
172 /
173 /
174 /
175 /
176 /
177 /
178 /
179 /
180 /
181 /
182 /
183 /
184 /
185 /
186 /
187 /
188 /
189 /
190 /
191 /
192 /
193 /
194 /
195 /
196 /
197 /
198 /
199 /
200 /
201 /
202 /
203 /
204 /
205 /
206 /
207 /
208 /
209 /
210 /
211 /
212 /
213 /
214 /
215 /
216 /
217 /
218 /
219 /
220 /
221 /
222 /
223 /
224 /
225 /
226 /
227 /
228 /
229 /
230 /
231 /
232 /
233 /
234 /
235 /
236 /
237 /
238 /
239 /
240 /
241 /
242 /
243 /
244 /
245 /
246 /
247 /
248 /
249 /
250 /
251 /
252 /
253 /
254 /
255 /
256 /
257 /
258 /
259 /
260 /
261 /
262 /
263 /
264 /
265 /
266 /
267 /
268 /
269 /
270 /
271 /
272 /
273 /
274 /
275 /
276 /
277 /
278 /
279 /
280 /
281 /
282 /
283 /
284 /
285 /
286 /
287 /
288 /
289 /
290 /
291 /
292 /
293 /
294 /
295 /
296 /
297 /
298 /
299 /
300 /
301 /
302 /
303 /
304 /
305 /
306 /
307 /
308 /
309 /
310 /
311 /
312 /
313 /
314 /
315 /
316 /
317 /
318 /
319 /
320 /
321 /
322 /
323 /
324 /
325 /
326 /
327 /
328 /
329 /
330 /
331 /
332 /
333 /
334 /
335 /
336 /
337 /
338 /
339 /
340 /
341 /
342 /
343 /
344 /
345 /
346 /
347 /
348 /
349 /
350 /
351 /
352 /
353 /
354 /
355 /
356 /
357 /
358 /
359 /
360 /
361 /
362 /
363 /
364 /
365 /
366 /
367 /
368 /
369 /
370 /
371 /
372 /
373 /
374 /
375 /
376 /
377 /
378 /
379 /
380 /
381 /
382 /
383 /
384 /
385 /
386 /
387 /
388 /
389 /
390 /
391 /
392 /
393 /
394 /
395 /
396 /
397 /
398 /
399 /
400 /
401 /
402 /
403 /
404 /
405 /
406 /
407 /
408 /
409 /
410 /
411 /
412 /
413 /
414 /
415 /
416 /
417 /
418 /
419 /
420 /
421 /
422 /
423 /
424 /
425 /
426 /
427 /
428 /
429 /
430 /
431 /
432 /
433 /
434 /
435 /
436 /
437 /
438 /
439 /
440 /
441 /
442 /
443 /
444 /
445 /
446 /
447 /
448 /
449 /
450 /
451 /
452 /
453 /
454 /
455 /
456 /
457 /
458 /
459 /
460 /
461 /
462 /
463 /
464 /
465 /
466 /
467 /
468 /
469 /
470 /
471 /
472 /
473 /
474 /
475 /
476 /
477 /
478 /
479 /
480 /
481 /
482 /
483 /
484 /
485 /
486 /
487 /
488 /
489 /
490 /
491 /
492 /
493 /
494 /
495 /
496 /
497 /
498 /
499 /
500 /
501 /
502 /
503 /
504 /
505 /
506 /
507 /
508 /
509 /
510 /
511 /
512 /
513 /
514 /
515 /
516 /
517 /
518 /
519 /
520 /
521 /
522 /
523 /
524 /
525 /
526 /
527 /
528 /
529 /
530 /
531 /
532 /
533 /
534 /
535 /
536 /
537 /
538 /
539 /
540 /
541 /
542 /
543 /
544 /
545 /
546 /
547 /
548 /
549 /
550 /
551 /
552 /
553 /
554 /
555 /
556 /
557 /
558 /
559 /
560 /
561 /
562 /
563 /
564 /
565 /
566 /
567 /
568 /
569 /
570 /
571 /
572 /
573 /
574 /
575 /
576 /
577 /
578 /
579 /
580 /
581 /
582 /
583 /
584 /
585 /
586 /
587 /
588 /
589 /
590 /
591 /
592 /
593 /
594 /
595 /
596 /
597 /
598 /
599 /
600 /
601 /
602 /
603 /
604 /
605 /
606 /
607 /
608 /
609 /
610 /
611 /
612 /
613 /
614 /
615 /
616 /
617 /
618 /
619 /
620 /
621 /
622 /
623 /
624 /
625 /
626 /
627 /
628 /
629 /
630 /
631 /
632 /
633 /
634 /
635 /
636 /
637 /
638 /
639 /
640 /
641 /
642 /
643 /
644 /
645 /
646 /
647 /
648 /
649 /
650 /
651 /
652 /
653 /
654 /
655 /
656 /
657 /
658 /
659 /
660 /
661 /
662 /
663 /
664 /
665 /
666 /
667 /
668 /
669 /
670 /
671 /
672 /
673 /
674 /
675 /
676 /
677 /
678 /
679 /
680 /
681 /
682 /
683 /
684 /
685 /
686 /
687 /
688 /
689 /
690 /
691 /
692 /
693 /
694 /
695 /
696 /
697 /
698 /
699 /
700 /
701 /
702 /
703 /
704 /
705 /
706 /
707 /
708 /
709 /
710 /
711 /
712 /
713 /
714 /
715 /
716 /
717 /
718 /
719 /
720 /
721 /
722 /
723 /
724 /
725 /
726 /
727 /
728 /
729 /
730 /
731 /
732 /
733 /
734 /
735 /
736 /
737 /
738 /
739 /
740 /
741 /
742 /
743 /
744 /
745 /
746 /
747 /
748 /
749 /
750 /
751 /
752 /
753 /
754 /
755 /
756 /
757 /
758 /
759 /
760 /
761 /
762 /
763 /
764 /
765 /
766 /
767 /
768 /
769 /
770 /
771 /
772 /
773 /
774 /
775 /
776 /
777 /
778 /
779 /
780 /
781 /
782 /
783 /
784 /
785 /
786 /
787 /
788 /
789 /
790 /
791 /
792 /
793 /
794 /
795 /
796 /
797 /
798 /
799 /
800 /
801 /
802 /
803 /
804 /
805 /
8
```

Query Match	10.9%	Score 76	DB 4	Length 557
Best Local Similarity	27.2%	Pred. No. 43		
Matches 31, Conservative	22	Mismatches 31	Indels 30	Gaps 7

```
Oy      4 IATKNFVSTTIS---KSPPA-----PEP-----RIOPS--EGENVGKGLIFSV 44  
       :|::: || |::||  
Db     295 VALIKHPMSQLSFDXLTKYFAQGSAPPPTWTFVELKNRGIEPMINMON---EGTGLEFSY 351
```

```
QY      45 NLTPPENVSQVTVPPYDEDYGGLVNTADASQSIIYQLVBEGKMKLKDHA 98
        :|:: : :|:: | :||| ||| |:| :
Db      352 DRTIPDLERARSFPDPK-----VRDLPFFRALEIKIVESGKEM-KEPS 397
```

RESULT 11
US-10-282-122A-68607

```

; Sequence 68607, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```

APPLICANT: Wang, Liangou
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELIITA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A

```

; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-10-282-122A-68607

```

Query Match	10.9%	Score	76;	DB	4;	Length	2358;
Best Local Similarity	22.3%	Pred. No.	3.3e+02;				
Matches	43;	Conservative	25;	Mismatches	55;	Indels	70;
						Gaps	9

Qy	2	TETATKNPVPVSTTIISKSPFAPEPRIOPSFGENVGKEGALLFS-----VNLTYP-----ENV	52
Db	1432	TTAAKKNFTIS-----NMOPHAGESITTYKALVDNHDNPVGCVPVAMSTNEG	146

QY	53	SQVTVPYVDE	DYGLG-----	RLVNTADASQSI	-----	RYQIVDE	-----	87
Db	1470	SQVETPLFTD	DGVAIVGLSR	SVGVKVSAL	ATGTYIAD	VHFLTHI	DEIDETMSEL	1522

```

QY -----KQKIML-----KDHGAETPNQOITFKALNT--SGEKKISPGIYNDQV-- 129
88 -----
1510 NPSOIIANGKDKMLITTFVKDKDGNIIIPROOVSGFSKXPTIKFQOAOQISPGREIERT 1588

```

QY 130 -----MVGYYN 136
 : | | |
DB 1590 TOSGTAOTGVN 1602

RESULT 12
US-10-282-122A-67315
; Sequence 67315, Application US/10282122A
; Publication No. US20040029129A1

```

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

US-10-282-122A-67315

Query Match	10.8%	Score 75.5	DB 4	Length 246
Best Local Similarity	21.4%	Pred. No. 16		
Matches 28, Conservative	25	Mismatches 49	Indels 29	Gaps 5

9 F E V S I I S K S F A P E K I - - Q F S F G E N V G A E G A U F S V N L I V P E N S Q V I V I F I D - - - - 0 2

Db 111 FGIGGLIDVASYSPLRVENQSRFGDTLGRYGV-----ETGSSYVMLPLXGPATP 159

Qy 63 -EDVGLRLVNTADASQSI-----YQVDEKKGKMLKDHAETPNQOITFKALN 112
: | : | : | : : | : | : | : : | : | :
Db 160 RQD--IGNLADTTYPMLSLGPMWLLKSGVQIIDRAKALDKDALLEGQSQDPYITFREAY 217

```
QY      113 YTSGEKKISPG 123
        : : | :: |
DB      218 FQNLERYRVKDQ 228
```

RESULT 13
US-10-282

Sequence 44691, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Heselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Esbe
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A

1 CURRENT FILING DATE: 2003-02-20
 2 PRIOR APPLICATION NUMBER: 60/191, 078
 3 PRIOR FILING DATE: 2000-03-21
 4 PRIOR APPLICATION NUMBER: 60/206, 848
 5 PRIOR FILING DATE: 2000-05-23

Query Match	10.7%	Score 75	DB 4	Length 640
Similarity	26.0%	Pred NO. 67		
Beet Local				
Matches 26	Conservative 19	Mismatches 47	Indels 8	Gaps 2

Qy	26	IQSPFENYNGKESALFFS	VNLTVPEKVSQV	-----	VYPYDDYGLGRVWNLADSQ	78
Db	64	VQGFLDQEVAKYLF	FTDSLKNGSEBVKLTSL	BEARVERCMAEKY	-RGSNNKKNSQ	12
Qy	79	SIITYQIVDEKGRKML	KDHGAETVTPQOITFR	KLANITSEK		118
Db	123	FTFDLITDDYQVL	VEKASDEBITQML	PLWRALSQ		162

RESULT 14
US-09-815-242-11293

Sequence 11293, Application US/09815242
Patent No. US20020061569A1

; GENERAL INFORMATION:

APPLICANT: Haselebeek, Karl R.
APPLICANT: Ohlsen, Robert
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T

APPLICANT: XU, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE NUMBER: PCT/US97/0111

FILE REFERENCE: BLIRKA.011A
CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

;
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

;
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11293
```

ORGANISM: *Helicobacter pylori*
US-09-815-242-11293

Query Match 10.7%; Score 75; DB 3; Length 736;
Best Local Similarity 19.7%; Pred. No. 82;
Matches 23; Conservative 29; Mismatches 63; Indels 2; Gaps 1.

RESULT 15
US-10-282

; Sequence 58624, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith

```



```

Sequence 362: Application US/10517939
Publication No. US20060003433A1
GENERAL INFORMATION:
APPLICANT: Steer, Brian
APPLICANT: Callen, Walter
APPLICANT: Healey, Shaun
APPLICANT: Hazlewood, Geoff
APPLICANT: Wu, Di
APPLICANT: Blum, David
APPLICANT: Etegegnatalan, Alireza
TITLE OF INVENTION: XLANAMES, NOCLETIC ACIDS ENCODING THEM
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
FILE REFERENCE: 56462007901
CURRENT APPLICATION NUMBER: US/10/517,939
CURRENT FILING DATE: 2004-12-13
PRIORITY APPLICATION NUMBER: PCT/US03/19153
PRIORITY FILING DATE: 2003-06-16
PRIORITY APPLICATION NUMBER: 60/389,299
PRIORITY FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 380
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 362
LENGTH: 1680
TYPE: FRT
ORGANISM: Unknown
FEATURES:
OTHER INFORMATION: Obtained from an environmental sample.
NAME/KEY: SIGNAL
LOCATION: (1)...(26)
US-10-517-939-362

```

Query Match	10.2%	Score 71.5;	DB 6;	Length 1680;
Best Local Similarity	21.3%	Pred. No. 84;		
Matches 37;	Conservative 27;	Mismatches 55;	Indels 55;	Gaps 8;

```
QY      I RTTEL-ATKAPVSTTISKSFFAPRPRIQPSRGEVNGEGALLFEVLN-TYPENASQTV 57
Db      1032 RTNVPAPATTTTIDVTLT-----RLQINTLLNAGGRLRIQGNAPTYITITIDDFV 1083
QY      58 YPVVDED-----YGLGR-----LVNTPAASQSIYQ---I 84
Db      1084 YQIDIDITAGLPRLPPQWFDLRPSLETFPYFGLNIYSTETLLMANETKRALFHHENV 1144
QY      85 VDEKGGKMLKDHGAELT-----PNOQITFKALNYSGEKISPGIYNQWVMGY 133
Db      1144 TAEKGHKRSSIAGRPNSFTVPEPBD-----FNFTDADRIYFAVENDIELYGH 1191
```

RESULT 7
US-11-194-246-282
; Sequence 282, Application US/11194246
; Publication No. US20050272089A1

```

1  TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND ME
2  TITLE OF INVENTION: USE
3  FILE REFERENCE: 00592.US1 (MAR 268.05920101)
4  CURRENT APPLICATION NUMBER: US/11/194,246
5  CURRENT FILING DATE: 2005-08-01
6  PRIOR APPLICATION NUMBER: US/10/274,586
7  PRIOR FILING DATE: 2002-10-21
8  PRIOR APPLICATION NUMBER: US 60/345,438
9  PRIOR FILING DATE: 2001-10-19
10 NUMBER OF SEQ ID NOS: 621
11 SOFTWARE: PatentIn version 3.0
12 SEQ ID NO 282
13 LENGTH: 357
14 TYPE: PRT
15 ORGANISM: HAEMOPHILUS INFLUENZAE
16 IS-11-194-246-282

```

	Query March	10.1k;	Score 70.5;	DB 7;	Length 357;	
	Best Local Similarity	25.8k;	Pred. No. 13;			
	Matches	32;	Conservative	23;	Mismatches	50;
					Indels	19;
					Gaps	7
Qy	16 SKSFADEPRIPQSFGF--NVGKGAALLSEVNLTPPENVSQTVVPYDDE-DYGAGRLV	71				
Dd	180 SNPFCCDESAAVEKFEDBLRELKKGSGSIGAKLTIVIAENVPEVGCEPFDRIADLALHNL	239				
Qy	72 NTADASGSIT---YQIVDEKGGKKMLDHGAEVTPNQGITPKALNTSGEKKISPGIYNDD	127				
Dd	240 MGINAAVGVGIVEIGDFAVVEQRG---SEHRDEMTPN--GFES-NHAGG---ILGGISSG	288				
Qy	128 QVMVV	131				
Dd	289 QPII	292				

RESULT 8
US-11-099-691-7

```

Sequence / Application ID: 06-099-691-7
Publication No. US20050260644A1

GENERAL INFORMATION:
APPLICANT: INCYTE PHARMACEUTICALS, 'INC.
APPLICANT: BANDMAN, Olga
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: YUE, Henry
APPLICANT: YANG, Y Tom
APPLICANT: PATTERSON, Chandra
APPLICANT: BAUGHN, Mariah R.
APPLICANT: YANG, Junming
TITLE OF INVENTION: CELL SIGNALING PROTEINNS
FILE REFERENCE: PF-0521 PCT
CURRENT APPLICATION NUMBER: US/11/099,691
CURRENT FILING DATE: 2005-04-06
PRIOR APPLICATION NUMBER: US/03/700,444
PRIOR FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 60/085,343
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/098,010
PRIOR FILING DATE: 1998-08-26
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PERL Program
SEQ ID NO 7
LENGTH: 898
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: Incyte Clone 2948818
US-11-099-691-7
```

Query Match	10.1%	Score 70.5	DB 7	Length 898
Best Local Similarity	23.2%	Pred. No. 45		
Matches	23	Conservative	26	Mismatches 27
				Indels 23
				Gaps 5

```

OY      28  PSCGENVGEGKALLTSVNLTVPEPNTSQTAVVPVYDEYDGLRLVNTADASQIIYO-----83
Db      351  PMSHVSVTGSSQSSVNSMGEVMEDESSSELYM--MHDE-----STINSSSVVHKKHV 401
OY      84  -LYDEKGMKLDHG---AEVTNPQOITFPKALNTYSGEK 118
Db      402  FIRDGAG-----HGDPREPRPTSVSQALHYNNRR 434

```

RESULT 9
US-11-124-367A-330
; Sequence 330, Application US/11124367M
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Bongjin Huang


```

RESULT 13
US-11-096-568A-6310
; Sequence 6310, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:

```

```

      APPLICANT: Alexandrov, Nickolai et al.
      TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
      FILE OF INVENTION: Thebydy
      FILE REFERENCE: 2750-1592PUS2
      CURRENT APPLICATION NUMBER: US/11/096,568A
      CURRENT FILING DATE: 2005-04-01
      NUMBER OF SEQ ID NOS: 34471
      SEQ ID NO 6310
      LENGTH: 450
      TYPE: PRT
      ORGANISM: Glycine max
      FEATURE:
      NAME/KEY: misc feature
      LOCATION: (1)_(450)
      OTHER INFORMATION: Cereb Seq. ID no. 14314718
      US-11-096-568A-6310

Query Match
Best Local Similarity 9.9%; Score 69.5; DB 7; Length 450;
Matches 22; Conservative 11; Mismatches 20; Indels 33; Gaps 3;

      35 GKEGALLFSVNLTVPENVSQVTVY-----PYVDE 63
      |||:::|||||:::|:|:|
      Db 361 GKEDFIYGVNLTF-VNLEANIYDGMNLNGOKPIMANCTFRGVSBDKGVLVLPAPEDDE 419
      |||:::|||||:::|:|:|

QY 64 DYGLGRLVNTADASQSIYQIVDEKG 89
      |||:::|||||:::|:|:|
      Db 420 DGGNGRIV-TVSLPRESLYOLDKLKG 444
      |||:::|||||:::|:|:|

RESULT 14
US-11-087-099-3301
Sequence 3301, Application US/11087099
Publication No. US20060041961A1
GENERAL INFORMATION:
APPLICANT: Abad, Mark S. et al.
TITLE OF INVENTION: Genes and Uses for Plant Improvement
FILE REFERENCE: 38-21(53450)B EP
CURRENT APPLICATION NUMBER: US/11/087,099
CURRENT FILING DATE: 2005-03-22
NUMBER OF SEQ ID NOS: 12464
SEQ ID NO 3301
LENGTH: 501
TYPE: PRT
ORGANISM: Staphylococcus aureus subsp. aureus MW2
US-11-087-099-3301

Query Match
Best Local Similarity 9.9%; Score 69; DB 7; Length 501;
Matches 33; Conservative 17; Mismatches 39; Indels 40; Gaps 7;

QY 7 KNFPVSTTISKSPFAPEDRIQPSF---GENV-----GKEGALLFSVNLTVPENVSQVTVYP 59
      |||:::|||||:::|:|:|
      Db 262 ENFMKAMDYIIIGGSGPLPSISQAFKQYGINNINGVGLTEAPLVVN--TPENSKR----- 314
      |||:::|||||:::|:|:|

QY 60 VYDEDYGLGLVNTADASQSIYQIVDEKGMKLDHGAETVTPNQ--QITFALNNTSGE 117
      |||:::|||||:::|:|:|
      Db 315 ---KPMISGKAVMFVDA-----RIIDNGEVEVPTGEIGELAIRA----- 350
      |||:::|||||:::|:|:|

QY 118 KGISPGIYN 126
      |||:::|||||:::|:|:|
      Db 351 KNVTTPGYWN 359
      |||:::|||||:::|:|:|

RESULT 15
US-10-793-626-2208
Sequence 2208, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PUS480US
CURRENT APPLICATION NUMBER: US/10/793,626

```

```

/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2208
/ LENGTH: 278
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
/ OTHER INFORMATION: amino acid sequence
US-10-793-626-2208

Query Match          9.8%; Score 68.5; DB 6; Length 278;
Best Local Similarity 30.4%; Pred. No. 15;
Matches 31; Conservative 12; Mismatches 48; Indels 11; Gaps 4

QY 3 EIAITKPNPVSSTISKSPFAEPRIQPSFGENVGKAGALLFSVNLTPVENVSQVTYYPYVD 62
   |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
Db 57 DIAAQVEGALVTLNKEDKA-EPALAKSFPKK--SNGGKTLTINIRKNAKWSNGDSVTAYD 113

QY 63 EDYGLGRVLNLTADASQSIITYQIVDEK-----GKKMLKDHG 97
   |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
Db 114 FVYAMRKVNPKTASE--FAYIMSDIKADEVNAAGKSKSYKDG 154

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 10, 2006, 09:55:20 / Search time 26.7 Seconds
(without alignments)
1734.477 Million cell updates/sec

Title: US-10-768-093-4
Sequence: 1 TTGACACATTACGAATGTTA.....TTGGTTACTACTGTAACCTAA 527

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-DB=ABSSWEB_pool/US10768093/runat_10042006_090301_4503/app_query.fasta.1
-UNITs=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs802p
-USER=us10768093 @GCN 1.1.348 @runat 10042006 090301 4503 -NCPU=6 -ICPU=3
-NO_MMAR -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.21.*
1: geneeqp19808.*
2: geneeqp19808.*
3: geneeqp20008.*
4: geneeqp20018.*
5: geneeqp20028.*
6: geneeqp20038.*
7: geneeqp20038.*
8: geneeqp20048.*
9: geneeqp20058.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	83.6	154	8	ADRI4912 Escherich
2	785	83.6	154	8	Adw47733 E. coli C
3	699	74.4	136	8	ADRI4916 CS6 relat
4	699	74.4	136	8	Adw47737 E. coli C
5	118	12.6	167	8	ADRI4913 Escherich
6	118	12.6	167	8	Adw47734 E. coli C
7	113.5	12.1	146	8	ADRI4917 CS6 relat
8	92.5	9.9	221	6	AAB37981 Human kin
9	90.5	9.6	268	9	ADW47927 Thermobif

10	87.5	9.3	911	6	ABR41065	ADRI41065 Human MAP
11	87.5	9.3	911	7	ADP68432	Adv68432 Human HIP
12	86	9.2	378	7	ADP95115	Adp95115 E. faeciu
13	85	9.1	506	4	ABR49208	Abd49208 V.cholera
14	85	9.1	541	6	ABU49234	Abu49234 Protein e
15	84.5	9.0	1239	9	ADRI86415	Adri86415 Aspergill
16	84	8.9	356	9	ABR40460	ABR40460 L. pneumo
17	84	8.9	384	9	ABR37118	ABR37118 L. pneumo
18	84	8.9	439	8	ADN19836	Adn19836 Bacterial
19	83	8.8	183	5	ABR49515	Abd49515 listeria
20	83	8.8	750	6	ADB09635	Adb09635 Allotococ
21	82	8.7	610	9	ABR37679	ABR37679 L. pneumo
22	82	8.7	610	9	ABR40990	ABR40990 L. pneumo
23	81	8.6	768	7	ADH88582	Adh88582 Enterococ
24	79.5	8.5	422	5	ADP28400	ADP28400 Bacterial
25	79.5	8.5	835	5	ABP26319	ABP26319 Streptoc
26	79.5	8.5	951	6	ABR41062	ABR41062 Human MAP
27	79.5	8.5	1171	4	ABR65661	ABR65661 Novel pro
28	79.5	8.5	1171	5	ABP69278	ABP69278 Human pol
29	79.5	8.5	1171	5	ABR16262	ABR16262 Human kin
30	79.5	8.5	1171	6	ABR41063	ABR41063 Human MAP
31	79.5	8.5	1171	8	ADP129268	ADP129268 Human PRO
32	79.5	8.5	1171	9	ADP18325	ADP18325 PRO polyp
33	79.5	8.5	1171	9	ADP18325	ADP18325 PRO polyp
34	79.5	8.5	1198	5	ABP69277	ABP69277 Human pol
35	78.5	8.4	332	3	ABG10781	ABG10781 Arabidops
36	78.5	8.4	355	3	ABG10780	ABG10780 Arabidops
37	78.5	8.4	355	3	ABR36463	ABR36463 A. thalia
38	78.5	8.4	367	3	ABG10779	ABG10779 Arabidops
39	78.5	8.4	989	9	ADV68430	Adv68430 Human hom
40	78.5	8.4	1191	5	ABO18296	ABO18296 Human hom
41	78.5	8.4	1198	5	ABG61938	ABG61938 Proteate
42	78.5	8.4	1198	5	ABO22555	ABO22555 Serine-ri
43	78.5	8.4	1198	6	ABG73319	ABG73319 Human kin
44	78.5	8.4	1198	6	ABR41064	ABR41064 Human MAP
45	78.5	8.4	1198	7	ADD18976	ADD18976 Human dis

ALIGNMENTS

RESULT 1	ADRI4912	standard; protein; 154 AA.
XX	ADRI4912;	
AC	ADRI4912;	
DT	04-NOV-2004 (first entry)	
XX		
DE	Escherichia coli E8775 antigen csaA protein.	
XX		
XX	CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter;	
KW	kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli;	
KW	Escherichia coli E8775.	
XX		
OS	Escherichia coli.	
XX		
PN	US2004156829-A1.	
XX		
PD	12-AUG-2004.	
XX		
PF	02-FEB-2004; 2004US-00768093.	
XX		
PR	13-MAY-1994; 94US-0023482.	
PR	24-JAN-1997; 97US-00788145.	
PR	10-JAN-2000; 2000US-00479877.	
XX		
PA	(USSA) US SEC OF ARMY.	
XX		
PI	Wolf MK, Cassels FJ, Boedeker EC,	
XX		
DR	WPI; 2004-592722/57.	
XX		
XX	N-PSDB; ADRI4911.	
XX		

Db 61 SerValaenLeuThrValProGluasnValSerGlnValThrValTyrProValTyrAsp 80
 Oy 303 GAAGATTAGGTTGGACGACTAGTAATAACCGCTGATGCTCCCAATCATTAATTCAC 362
 Db 81 GluAspTyrGlyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIleIleTyr 100
 Oy 363 CAGATTGTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGTCAGAGGTTACACT 422
 Db 101 GlnIleValaIAspGluIleGlyLysLysMetLeuLysAspHisGlyAlaGluValThrPro 120
 Oy 423 AATCAACAATAATTAATCTTTAAACGCTGAATTAATTAAGGGGAAAAAATATCTCT 482
 Db 121 AsnGlnGlnIleThrPheLysAlaLeuAsnTyrThrSerGlyGluLysIleSerPro 140
 Oy 483 GGAATATATAACGATCAGGTTATGTTGGTTACTACTGTAAC 524
 Db 141 GlyIleTyrAsnAspGlnValMetValGlyTyrTyrValAsn 154

RESULT 3

ADRI4916 standard; protein; 136 AA.

ADRI4916,

04-NOV-2004 (first entry)

CS6 related protein #1.

CS6, csaA, csaB, csaC, csaD; origin of replication, lac promoter;

kanamycin resistance gene; CS6a; CS6b; enterotoxigenic E. coli.

Unidentified.

US2004156829-A1.

12-AUG-2004.

02-FEB-2004; 2004US-00768093.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(USSA) US SEC OF ARMY.

Wolf MK, Casseels FJ, Boedeker EC;

WPI; 2004-592722/57.

Inducing the production of antibodies against CS6 protein, useful for preventing pathological effects of enterotoxigenic E. coli, by administering a composition comprising transformed bacteria producing CS6 antigens.

Claim 4; SEQ ID NO 9; 22pp; English.

The invention relates to a method of inducing, in a susceptible host, the production of antibodies against a CS6 protein, comprising administering a composition of matter comprising bacteria transformed with a plasmid containing genes csaA, csaB, csaC and csaD; an origin of replication, a lac promoter and a kanamycin resistance gene, where the bacteria expresses both CS6A and CS6B proteins. The invention also relates to a composition of matter comprising a protein in a pharmaceutical carrier. The pharmaceutical carrier is a carbonated liquid. The method is useful for inducing the production of antibodies against the CS6 protein. The protein and composition are used as vaccines for preventing pathological effects of enterotoxigenic E. coli. This sequence represents a CS6 related protein of the invention.

Sequence 136 AA;

Alignment Scores:

Pred. No.: 2.92e-82 Length: 136
 Score: 699.00 Matches: 136
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 74.4% Indels: 0
 DB: 8 Gaps: 0

US-10-768-093-4 (1-527) x ADRI4916 (1-136)

Oy 117 AGAACAAGATAGGACATAAACTCCAGTATCAACGATATTCAAAAAGTTT 176
 Db 1 ArgThrGlnIleAlaThrLysAsnPheProValSerThrTrpIleSerLysSerPhe 20
 Oy 177 GCACCTGAACCAAGAAATACGCTCTTTTGGTGAATAATGTTGAAAGAAAGACCTTTA 236
 Db 21 AlaProGluProArgLysGlnProSerPheGlyGluAsnValGlyLysGlyAlaLeu 40
 Oy 237 TTTATTAGTGAAGCTTAACCTGTTCCGAAATGATCCAGGTAAGGCTACCGCTGT 296
 Db 41 LeuPheSerValaenLeuThrValProGluasnValSerGlnValThrValTyrProVal 60
 Oy 297 TATGATGAAGATTATGGGTTAGACGACTAGTAATAACCGCTGATGCTCCCAATCAATA 356
 Db 61 TyrAspGluAspTyrGlyLeuGlyArgLeuValaenThrAlaAspAlaSerGlnSerIle 80
 Oy 357 ATCTACAGATTGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGTCAGAGGTT 416
 Db 81 IleTyrGlnIleValaIAspGluIleGlyLysLysMetLeuLysAspHisGlyAlaGluVal 100
 Oy 417 ACACCTAATCAACAATAATTAATTAACCGCTGAATTAATTAAGGGGAAAAAATA 476
 Db 101 ThrProAsnGlnIleThrPheLysAlaLeuAsnTyrThrSerGlyGluLysIle 120
 Oy 477 TCTCCTGAATATATAACGATCAGGTTATGTTGGTTACTACTGTAAC 524
 Db 121 SerProGlyIleTyrAsnAspGlnValMetValGlyTyrTyrValAsn 136

RESULT 4

ADM47737 standard; protein; 136 AA.

ADM47737,

21-APR-2005 (first entry)

E. coli CS6 csaA mature protein.

antibacterial; vaccine; DNA purification; colonization factor antigen.

Escherichia coli.

US2005025787-A1.

03-FEB-2005.

12-JAN-2004; 2004US-00754641.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(WOLF/) WOLF M K.

(CAS/) CASSELS F J.

(BOED/) BOEDEKER E C.

Wolf MK, Casseels FJ, Boedeker EC;

WPI; 2005-131784/14.

Inducing in a susceptible host, the production of antibodies against CS6 protein, useful for stimulating protective antibodies against Escherichia coli by administering a composition comprising bacteria transformed with a plasmid.

XX PS Claim 4; SEQ ID NO 9; 23pp; English.

CC The invention relates to a method of inducing in a susceptible host, the

CC production of antibodies against CS6 protein comprising giving a

CC composition of matter made of bacteria transformed with a plasmid, which

CC contain genes *csaA* and *csaB*, all of *csaC* and DNA sequence *csaD* that

CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of

CC replication, a *lac* promoter, and a kanamycin resistance gene, where the

CC bacteria overexpress both *CS6A* and *CS6B* proteins. The method and proteins

CC are useful for stimulating protective antibodies against enterotoxigenic

CC *Escherichia coli*. This sequence corresponds to the *E. coli* *CS6 csaA*

CC mature protein

CC

XX SQ Sequence 136 AA;

Alignment Scores:

Pred. No.:	2,926-82	Length:	136
Score:	699.00	Matches:	136
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	74.4%	Indels:	0
DB:	9	Gaps:	0

US-10-768-093-4 (1-527) x ADW47737 (1-136)

QY 117 AGAACAAGAAATAGCGACTAAACCTCCGATCAACGACTATTTCAAAAAGTTT 176

DB 1 ArgThGluIleAlaThrIleHisAsnProValSerThrIleSerIleSerPhe 20

QY 177 GCACCTGAACCAACGACTACGCTTTCTTTGGTGAAGAAAGGAGGACTTTA 236

DB 21 AlaProGluProArgIleGlnProSerPheGlyGluAsnValGlyGluGlyAlaLeu 40

QY 237 TATTATTAGTGAAGTAACTGCTCTGAAAGTATCCAGGTACGGCTACCTGTT 296

DB 41 LeuPheSerValAsnLeuThrValProGluAsnValSerGlnValThrValIleProVal 60

QY 297 TATGATGAAGATTATGCTTAGACGACACTAGTAATATACCGCTGATCTTCCCATCAATA 356

DB 61 TyrAspGluAspTyrGlyLeuGlyArgLeuValAsnThrAlaAspIleSerGlnSerIle 80

QY 357 ATCTACCAATGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGCTGCAGGCTT 416

DB 81 IleTyrGlnIleValAlaAspGluIleGlySerIleGlySerIleGlyAlaGluVal 100

QY 417 AACACCTAATCAACAATACTTTAAAGCGCTGATTAATATGCGGGGAAAAAATA 476

DB 101 ThrProAsnGlnGlnIleThrPheIleValAlaLeuAsnThrIleSerGlyGluIleVal 120

QY 477 TCTCCTGAATATATTAACGATCAGTTAGTTGCTTACTATGTAAC 524

DB 121 SerProGlyIleTyrAsnAspGlnValMetValGlyTyrTyrValAsn 136

RESULT 5

ID ADR14913 standard; protein; 167 AA.

XX ADR14913,

XX 04-NOV-2004 (first entry)

DE *Escherichia coli* E8775 antigen *csaB* protein.

KM CS6; *csaA*; *csaB*; *csaC*; *csaD*; origin of replication; *lac* promoter;

KM kanamycin resistance gene; *CS6A*; *CS6B*; enterotoxigenic *E. coli*;

KM *Escherichia coli* E8775.

OS *Escherichia coli*.

XX US2004156829-A1.

XX 12-AUG-2004.

XX PF 02-FEB-2004; 2004US-00768093.

XX PR 13-MAY-1994; 94US-00243482.

PR 24-JAN-1997; 97US-00788145.

PR 10-JAN-2000; 2000US-00479877.

XX (USSA) US SEC OF ARMY.

PI Wolf MK, Casele PJ, Boedeker EC,

XX WPI; 2004-592722/57.

DR

XX Inducing the production of antibodies against CS6 protein, useful for

PT preventing pathological effects of enterotoxigenic *E. coli*, by

PT administering a composition comprising transformed bacteria producing CS6

XX antigens.

XX Disclosure; SEQ ID NO 6; 22pp; English.

PS The invention relates to a method of inducing, in a susceptible host, the

XX production of antibodies against a CS6 protein, comprising administering

CC a composition of matter comprising bacteria transformed with a plasmid

CC containing genes *csaA*, *csaB*, *csaC* and *csaD*, an origin of replication, a

CC *lac* promoter and a kanamycin resistance gene, where the bacteria

CC express both *CS6A* and *CS6B* proteins. The invention also relates to a

CC composition of matter comprising a protein in a pharmaceutical carrier.

CC The pharmaceutical carrier is a carbonated liquid. The method is useful

CC for inducing the production of antibodies against the CS6 protein. The

CC protein and composition are used as vaccines for preventing pathological

CC effects of enterotoxigenic *E. coli*. This sequence represents the

CC *Escherichia coli* E8775 antigen *csaB* of the invention.

XX SQ Sequence 167 AA;

Alignment Scores:

Pred. No.:	7,746-06	Length:	167
Score:	118.00	Matches:	49
Percent Similarity:	42.1%	Conservative:	26
Best Local Similarity:	27.5%	Mismatches:	67
Query Match:	12.6%	Indels:	36
DB:	8	Gaps:	9

US-10-768-093-4 (1-527) x ADR14913 (1-167)

QY 63 ATGAAGAAACAAATGTTTAAATCTTCTGCTTATTCGAGCCATGCCAACA 122

DB 2 LeuIleValIleIleSerAlaIleLeuIleAla-----GlyThrSerGlyValVal 19

QY 123 GAATATGCGACT-----AAAACCTCCGATCAACGACTATTTCAAAAAGTTT 173

DB 20 AsnAlaGlyAsnTyrGlnTyrIleSerLeuValAsnValAsnIleGlnIleAsnPro 39

QY 174 TTGACCTGAA-----CCAGATATACGCTCTTTGTTGGAATAAGTTGA 221

DB 40 Ile---ProAspIleAspSerAlaValArgIleIleProValAsnTyrAspSerPro 58

QY 222 AAGGAAGAGCTTTATTTATTTAGTGAATTAATGCTTCTGAAATATATATCCAGGTA 281

DB 59 LysLeuAspSerGlnLeuTyrThrValGluMetThrIleProAlaGlyValSerAlaVal 78

QY 282 ACGGCTACCT-----GTTATGATGAAGATTATGGTTAGACGACTAGTAAT 332

DB 79 LysIleAlaProThrAspSerLeuThrSerSerGlyGlnIleGlyLysLeuValAsn 98

QY 333 ACCGCTGATGCTTCCCAATCAATATATCAACGATGTTGATGAGAAAGGAAAAAATG 392

DB 99 ValAsnAsnProAspGlnAsnMetCasnTyrTyrIle-----GAGGTTACA 419

QY 393 TTAAGAATCATGCTGCA-----GAGGTTACA 419

DB 111 ArgIleAspSerGlyAlaGlyAsnPheMetAlaGlyGlnIleGlySerPheProValLys 130

OY 420 CCTAATCAACAAATTAACCTTTAAAGCGCTGAATTAATCTAGCGGGAA-----AAA 470
 DB 131 GluAenThrseryTyrThrseryAlaile---TyrThrglyGlyGlyProAsnser 149
 OY 471 AAAATATCTCTCGAATATATACATCAGCTTATGTTGTTACTATGTAAC 524
 DB 150 GlyTyrSerSerglyThrTyralaGlyAsnLeuthrValSerPheTySerAsn 167
 RESULT 6
 ADM47734
 ID ADM47734 standard; protein; 167 AA.
 AC ADM47734;
 XX 21-APR-2005 (first entry)
 DE E. coli CS6 ccsB full length protein.
 KM antibacterial; vaccine; DNA purification; colonization factor antigen.
 XX Escherichia coli.
 OS US2005025787-A1.
 PN 03-FEB-2005.
 PF 12-JAN-2004; 2004US-00754641.
 PR 13-MAY-1994; 94US-00243482.
 PR 24-JAN-1997; 97US-00788145.
 PR 10-JAN-2000; 2000US-00479877.
 PA (WOLF/) WOLF M. K.
 PA (CASS/) CASSELS F. J.
 PA (BOED/) BOEDEKER E. C.
 PI Wolf MK, Casseels FJ, Boedeker EC;
 DR WPI; 2005-131784/14.
 PT Inducing in a susceptible host, the production of antibodies against CS6
 PT protein, useful for stimulating protective antibodies against Escherichia
 PT coli by administering a composition comprising bacteria transformed with
 PT a plasmid.
 PS Disclosure; SEQ ID NO 6; 23pp; English.
 XX The invention relates to a method of inducing in a susceptible host, the
 CC production of antibodies against CS6 protein comprising giving a
 CC composition of matter made of bacteria transformed with a plasmid, which
 CC contain genes csaA and csaB, all of csaC and DNA sequence csaD that
 CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of
 CC replication, a lac promoter, and a kanamycin resistance gene, where the
 CC bacteria overexpress both CS6A and CS6B proteins. The method and proteins
 CC are useful for stimulating protective antibodies against enterotoxigenic
 CC Escherichia coli. This sequence corresponds to the E. coli CS6 ccsB
 CC protein
 XX
 XX Sequence 167 AA;
 Alignment Scores:
 Pred. No.: 7.74e-06 Length: 167
 Score: 118.00 Matches: 49
 Percent Similarity: 42.14 Conservative: 26
 Best Local Similarity: 27.54 Mismatches: 67
 Query Match: 12.64 Indels: 36
 DB: 9 Gaps: 9
 US-10-768-093-4 (1-527). x ADM47734 (1-167).
 OY 63 ATGAAGAAACAAATGTTTATCTGCTTCAATTCGCGAGCCATCGACGAACA 122
 DB 2 LeuylalyslellelserAlaileAlaLeuAla-----GlyThrseryAlaVal 19

OY 123 GAATAGGACT-----AAAACCTCCAGTATACAGCATATTTCAAAAATT 173
 DB 20 AsnAlaGlyAenTpgIntyTyLyseryLeuApyValAenValAenlleGluGlnAsnDhe 39
 OY 174 TTTGCACCTGAA-----CCACGAATACAGCTTCTTTTGGTGAATGTTGGA 221
 DB 40 Ile---ProAepIleApseryAlaValArgllelleProValAenTyraPserAspPro 58
 OY 222 AAGGAAGAGCTTATTTATTTAGTGAACCTTAACCTGTTCCGAAATGATCCAGGTA 281
 DB 59 LysleuApserySergInleuthrThrValGluMetThrIleProAlaGlyValSerAlaVal 78
 OY 282 ACGGCTTACCT-----GTTTATGATGAAGATTAAGGTTAGACGACTGTAAT 332
 DB 79, LysIleAlaProThraPserLeuthrSerseryGlyGlnIlelleGlyLeuValAsn 98
 OY 333 ACCGCTATGCTTCCCATCAATATCTACAGATTTGTGTGATGAAGAAAGGAAAAATG 392
 DB 99 ValAsnAenProAepGlnAenMetAsnTyTyIle-----GAGTTTACA 419
 OY 393 TTTAAAGATCATGGTGCA-----GAGTTTACA 419
 DB 111 ArglyApseryGlyAlaGlyAenPheMetAlaGlyGlnLysGlySerPheProValLys 130
 OY 420 CCTAATCAACAAATTAACCTTTAAAGCGCTGAATTAATCTAGCGGGAA-----AAA 470
 DB 131 GluAenThrseryTyrThrseryAlaile---TyrThrglyGlyGlyProAsnser 149
 OY 471 AAAATATCTCTCGAATATATACATCAGCTTATGTTGTTACTATGTAAC 524
 DB 150 GlyTyrSerSerglyThrTyralaGlyAsnLeuthrValSerPheTySerAsn 167
 RESULT 7
 ADM4917
 ID ADM4917 standard; protein; 146 AA.
 AC ADM4917;
 XX 04-NOV-2004 (first entry)
 DE CS6 related protein #2.
 XX CS6; csaA, csaB, csaC, csaD, origin of replication; lac promoter;
 KM kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.
 OS Unidentified.
 XX US2004156829-A1.
 PN 12-AUG-2004.
 PD 02-FEB-2004; 2004US-00768093.
 PR 13-MAY-1994; 94US-00243482.
 PR 24-JAN-1997; 97US-00788145.
 PR 10-JAN-2000; 2000US-00479877.
 PA (USSA) US SEC OF ARMY.
 PI Wolf MK, Casseels FJ, Boedeker EC;
 DR WPI; 2004-592722/57.
 PT Inducing the production of antibodies against CS6 protein, useful for
 PT preventing pathological effects of enterotoxigenic E. coli, by
 PT administering a composition comprising transformed bacteria producing CS6
 PT antigens.
 PS Claim 6; SEQ ID NO 10; 22pp; English.
 CC The invention relates to a method of inducing, in a susceptible host, the
 CC production of antibodies against a CS6 protein, comprising administering

CC a composition of matter comprising bacteria transformed with a plasmid
 CC containing genes *csbA*, *csbB*, *csbC* and *csd*, an origin of replication, a
 CC lac promoter and a kanamycin resistance gene, where the bacteria
 CC expresses both *CS6A* and *CS6B* proteins. The invention also relates to a
 CC composition of matter comprising a protein in a pharmaceutical carrier.
 CC The pharmaceutical carrier is a carbonated liquid. The method is useful
 CC for inducing the production of antibodies against the *CS6* protein. The
 CC protein and composition are used as vaccines for preventing pathological
 CC effects of enterotoxigenic *E. coli*. This sequence represents a *CS6*
 CC related protein of the invention.

XX Sequence 146 AA;

Alignment Scores:
 Pred. No.: 2,84e-05 Length: 146
 Score: 113.50 Matches: 43
 Percent Similarity: 43.0% Conservative: 22
 Best Local Similarity: 28.5% Mismatches: 55
 Query Match: 12.1% Indels: 31
 DB: 8 Gaps: 7

US-10-768-093-4 (1-527) x ADR14917 (1-146)

QY 135 AAAAATCTCCAGTATCAACGACATATTGCAAAAGTTTTCGACCTGAA----- 185
 DB 6 LysSerLeuAspValAsnValAsnIleGluGlnAsnPhile--ProAspIleAspSer 24
 QY 186 ---CCACGCAATACAGCCTTCTTTGGTGAATAATGTTGAAGAGAGACCTTATATTT 242
 DB 25 AAlaValArgIleIleProValAsnTyrAspSerAspProValLeuAspSerGlnLeuTyr 44
 QY 243 AGTGTGAATTAACGTGTTCTCTGAAAAATGATATCCACAGTAACGCTTACCT----- 293
 DB 45 ThrValGluMetThrIleProAlaGlyValSerAlaValIleAlaProThrAspSer 64
 QY 294 GTTATGATGAAGATTATGGTAGCAGCATAGTAATACCGCATGCTTCCCAATCA 353
 DB 65 LeuThrSerSerGlyGlnGlnIleGlyValLeuValAsnValAsnAsnProAspGlnAsn 84
 QY 354 ATATATCTACCAAGATTGTTCATGATGAAGAGAGAAAAATGTTAAAGATCATGTGCA-- 410
 DB 85 MetAsnTyrTyrIle-----ArgLysAspSerGlyAlaGly 96
 QY 411 -----GAGGTACACCTTAATCAACAAATAACTTT 440
 DB 97 AsnPhenMetAlaGlyGlnLysGlySerPheProValLysGlnAsnThrSerTyrThrPhe 116
 QY 441 AAAAGCGTGAATTAATTAACAGCGGAGAA-----AAAAAATATCTCTGGAATATAT 491
 DB 117 SerAlaIle--TyrThrGlyGlyGlnTyrProAsnSerGlyTyrSerSerGlyThrTyr 135
 QY 492 AACGATCAGATTATGTTGGTTACTATGTAAC 524
 DB 136 AlaGlyAsnLeuThrValSerPheTyrSerAsn 146

RESULT 8
 ID AAE37981 standard; protein; 221 AA.
 XX AAE37981,
 AC AAE37981,
 XX 06-NOV-2003 (first entry)
 DT XX
 DE Human kinase and phosphatase (KPP-26) protein.
 XX
 XX Human; kinase; phosphatase; KPP; cell proliferative disorder; hepatitis;
 KW atherosclerosis; cirrhosis; haemoglobinuria; polycythaemia vera; cancer;
 KW psoriasis; thrombocytopenia; developmental disorder; Reiter's syndrome;
 KW renal tubular acidosis; anaemia; mental retardation; Alzheimer's disease;
 KW neurological disorder; Parkinson's disease; epilepsy; glomerulonephritis;
 KW autoimmune disorder; inflammatory disorder; AIDS; Goodpasture's syndrome;
 KW acquired immune deficiency syndrome; immunosuppressive; Crohn's disease;
 KW noctropic; transgenic; dermatitis; multiple sclerosis; diabetes mellitus;

KW allergy; gout; Grave's disease; Hashimoto's thyroiditis; bowel syndrome;
 KW osteoporosis; rheumatoid arthritis; Sjogren's syndrome; ophthalmological;
 KW gene therapy; asthma; anticonvulsant; uteropathic; pancreatitis.

XX Homo sapiens.

OS WO2003050084-A2.

XX 19-JUN-2003..

XX 06-DEC-2002; 2002MO-US039126.

XX 07-DEC-2001; 2001US-0340235P.

XX 19-DEC-2001; 2001US-0343007P.

XX 21-DEC-2001; 2001US-0343546P.

XX 04-FEB-2002; 2002US-0354388P.

XX 15-FEB-2002; 2002US-0357675P.

XX (INCY-) INCYTE GENOMICS INC.

XX Kable AE, Chien D, Wilson AD, Swarnakar A, Gorvad AE;
 PI Hafalia AJA, Emerling BM, Ramkumar J, Jin P, Griffin JA, Margulis JP;
 PI Baughn MR, Chawla NK, Lehr-Mason PM, Khare R, Lee S, Hawkins PR;
 PI Becha SD, Lee SY, Sprague WW, Zebardjadian Y;

XX WPI, 2003-532894/50.

XX N-PSDB; AAD57353.

XX New human kinases and phosphatases and polynucleotides, useful for
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
 PT cancer or hepatitis.

XX Claim 1; Page 230; 282pp; English.

XX The invention relates to an isolated polypeptide, which is a human kinase
 CC and phosphatase (KPP). KPP agonists and antagonists are useful for
 CC diagnosing, treating or preventing disorders associated with aberrant
 CC expression of KPP, particularly cell proliferative disorders (e.g.
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
 CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC thrombocytopenia or cancer), developmental disorders (eg. renal tubular
 CC acidosis, anaemia or mental retardation), neurological disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease or epilepsy), autoimmune/
 CC inflammatory disorders (e.g. AIDS; acquired immune deficiency syndrome,
 CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's
 CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,
 CC gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome,
 CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's
 CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,
 CC bacterial, fungal, parasitic, protozoan or helminthic infections. The KPP
 CC is useful in assessing the effects of exogenous compounds on the
 CC expression of nucleic acids and kinases and phosphatases. KPP gene is
 CC useful in gene therapy and for creating transgenic animals to model human
 CC disease. The present sequence is human KPP protein

XX Sequence 221 AA;

Alignment Scores:
 Pred. No.: 0.0198 Length: 221
 Score: 92.50 Matches: 41
 Percent Similarity: 42.2% Conservative: 27
 Best Local Similarity: 25.5% Mismatches: 72
 Query Match: 9.9% Indels: 21
 DB: 6 Gaps: 7

US-10-768-093-4 (1-527) x AAE37981 (1-221)

QY 93 CTGCTGATTCGGCAGCGCAACGCAAGAAATAGCAATTAATAACTTCCAGTATCA 152
 DB 41 MetThrGlyTyrGlySerHisSerLysValTyrSerGlnSerLysValAsnIleProLeuSer 60
 QY 153 -----ACGACATTTCAAAAAGTTTTCGACCTGAGACGCAATACAG--- 197


```

Db      61 GlnProAlaThrThrValSerThrSerLeuProValProAsnProSerLeuProTyr 80
Qy      198 -----CCTCTTTGGTGAAGTGTGAAGAGAGAGCTTTATTA 239
Db      81 GluGlnThrIleValPheProGlySerThrGlyHisIleValIleThrSerAlaSer 100
Qy      240 TTTACTGTGAAC-----TTAACTGTCCTGAAATGTA-----TCCGAGGTA 281
Db      101 ThrSerValThrGlyGlnValLeuGlyGlyProHisAsnLeuMetArgArgSerThrVal 120
Qy      282 ACGGTCTACCTGTTTATGATGAAGATTATGGTTGAGCGACTGTAATACCGCTGAT 341
Db      121 SerLeuLeuAspThrTyr---GlnLysCysGlyLeuLysArgLysSerGlnGlnIleGln 139
Qy      342 GCTTCCCAATCAATATATCAACAGATGTTGATGAGAAAGGAAAAAATGTTAAAGAT 401
Db      140 AsnThrSerSerVal-----GlnIleIleGlnGlnHis---ProProMetIleGlnHis 156
Qy      402 CATGGTGAAGAGGTTACACCTTAATCAACAATTAATCTTTAAAGCGCTGATTAATAC 461
Db      157 AsnAlaSerGlyAlaThrValAlaThrAlaThrThrSerThrAlaThrSerLysAsnSer 176
Qy      462 GGGGAAAAAAATATCTCTCGAATATATTAACGATCAGCTTATGGTTGTTACTATGTA 521
Db      177 GlySerProThrIleHisProSerGlnTyrProAlaGlnPheAlaHisGlnThrTyrIle 196
Qy      522 AAC 524
Db      197 Ser 197

RESULT 9
ADM47927
ID      ADM47927 standard; protein; 268 AA.
XX      ADM47927;
XX      DT
XX      07-APR-2005 (first entry)
XX      Thermobifida fusca protein kinase C-terminal domain.
XX      Protein Kinase; antibacterial; tuberculoestatic.
XX      OS
XX      Thermobifida fusca.
XX      PN
XX      MO2005007880-A1.
XX      PD
XX      27-JAN-2005.
XX      PF
XX      19-JUL-2004; 2004MO-IB003096.
XX      PR
XX      18-JUL-2003; 2003US-0487943P.
XX      PA
XX      (INSP ) INST PASTEUR.
XX      PI
XX      Alzari P, Boitel B, Villarino A, Fernandez P, Cole S;
XX      MPI; 2005-112885/12.
XX      PT
XX      Identifying substances that modulates pknB protein kinase or pspP2
XX      phosphatase activity, for identifying antibacterial substances, by
XX      comparing the activity from cells contacted with the substance to cells
XX      not contacted with the substance.
XX      Example: SEQ ID NO 6; 64bp; English.
XX      The present sequence is that of the C-terminal domain of a Ser/Thr
XX      protein kinase of Thermobifida fusca. Sequence homology is shown to the C
XX      -terminal domain of Ser/Thr protein kinase PknB ADM47924 of Mycobacterium
XX      tuberculosis. M. tuberculosis PknB, PknA and protein phosphatase PspP
XX      ADM47922, along with other signaling modulators, co-ordinately regulate
XX      cell elongation during growth. The invention provides claimed methods for
XX      identifying substances that modulate the activity of M. tuberculosis PknB

```

```

CC      or PspP (or related sequences). Such substances are potential
CC      antibacterial agents.
XX      SQ
XX      Sequence 268 AA;
XX      Alignment Scores:
XX      Pred. No.: 0.0396 Length: 268
XX      Score: 90.50 Matches: 49
XX      Percent Similarity: 42.1% Conservative: 28
XX      Best Local Similarity: 26.8% Mismatches: 54
XX      Query Match: 9.6% Indels: 52
XX      Db: 9 Gaps: 10
XX      US-10-768-093-4 (1-527) x ADM47927 (1-268)
Qy      69 AAACAATGGTTTATATCTAAT-----CTTGCTTCATTCGAGACCATCC 116
Db      91 LysGluValIGlyLeuThrAspIleValGlnAspGlnIleThrSerPheAspAsnProPro 110
Qy      117 AGAACGAAATAGCGACTTAAACTTCCA-----GTA 149
Db      111 GlyThrValIleThrThrIleProAlaProGlyGlnLysAlaAsnArgGlnGluSerVal 130
Qy      150 TCAACGACTATTTCAAAAGTTTTCACACTGAAACGACAAATACACCTCTTTGGT 209
Db      131 ThrLeuThrIleSerValGlyPhe-----ProMetProAsnVal-----ValGly 145
Qy      210 GAAATGTTGAAAGAGAGAGCTTATTTATTTAGTGAACCTTAACCTTCTCGAAAT 269
Db      146 GlnLysValIleAspPheAlaArgArgLeuGlnSerSerAspLeu----- 160
Qy      270 GTATCCAGGTACCGCTTACCTGTTTATGATGAAGATTATGGTTAGACGACTAGTA 329
Db      161 -----GluValThrValIleGlnGlnHisAspGlnValProGlnGlyHisValIle 178
Qy      330 AAT-----ACCGCTGATGCTTCCCAATCAATA----- 356
Db      179 SerGlnGlnProGlnValAlaGlnThrThrValGlnGlnSerValThrLeuThrVal 198
Qy      357 -----ATCTACCAAGATTGTTGAT-----GAGAAAGGG 383
Db      199 SerSerGlyProGlnLeuValGlnValProAspIleArgLysTyrTrpLysValAspLysAla 218
Qy      384 AAAAAAATGTTAAAGATCATGTCGACAGGTTTACCTTAATCAACAAATACTTTTAA 443
Db      219 ArgLysGlnLeuGlnGlnArgGlyPheGlnValThrValHisGlnValIleGlyAsnArg 238
Qy      444 CGCGTAATTATACT--AGCGGGGAAAAAAATATCTCGTGAATATATACGATGAG 500
Db      239 ValGlyAspTyrAsnProLysGlyGln-----AlaPro-LysGlySerThrIleGln 255
Qy      501 GTTATGG 507
Db      255 uilletrp 257

RESULT 10
ABR41065
ID      ABR41065 standard; protein; 911 AA.
XX      ABR41065;
XX      DT
XX      22-MAY-2003 (first entry)
XX      DE
XX      Human MAP kinase cascade activator #50.
XX      Human, Euk1 phosphorylation; Euk1 phosphorylation kinase; virulence;
XX      antiinflammatory; immunomodulator; cytotoxic; antiallergic; anti-HIV;
XX      antitubercular; antiparasitic; antidiabetic; antineoplastic; gene therapy;
XX      inflammation; autoimmune disease; viral disease; cancer; diabetes;
XX      rheumatoid arthritis; asthma; allergic rhinitis; AIDS; viral hepatitis;
XX      Iga nephritis.
XX      Homo sapiens.
XX      OS

```


Db 74 GIUGINRILLEVALPHEPROGLYSERTHRGLYHISILEVALTHRSERLASER 93
 QY 240 TTTAGTGTGAAC-----TTACTGTCCTGAAATATGTA-----TCCACAGTA 281
 Db 94 THRSERVALTHRGILGVALLEUGLYGPROHLSAENLEMECARGRARSERTHRYAL 113
 QY 282 ACGGCTTACCCCTGTTATGATGAGATTATGGGTTAGACAGCTAATAATCCGCTGAT 341
 Db 114 SERLEULEAAPTHTYR---GlnLysCysGLYLeuLYARGLYserGLUileGLU 132
 QY 342 GCTTCCCATCATATATCTACACGATTTGTGATGAGAAAAGGAAAAATGTTAAAAAGT 401
 Db 133 ASNTHRSERSERVAL-----GlnIleIleGLUgluHis---ProPrometIleGlnsn 149
 QY 402 CATGTGACAGAGCTTACACCTAATCAACAATATCACTTTAAAGCGCTGATTAATACTAGC 461
 Db 150 AENALASERGLYALATHRYALATHRYALATHRYALATHRYALATHRYALATHRYAL 169
 QY 462 GGGGAAAAAAATATCTCTCGAATATAT 491
 Db 170 GLYSERSEN-----SerGLUglYserPTyr 177
 RESULT 12
 ID ADC95115 standard; protein, 378 AA.
 XX ADC95115,
 AC
 DT 01-JAN-2004 (first entry)
 XX
 DE E. faecium protein sequence SEQ ID 4742.
 XX
 KM Vaccine; urinary tract infection; bacteremia; endocarditis; wound;
 abdominal-pelvic infection.
 XX
 OS Enterococcus faecium.
 FN US6583275-B1.
 XX 24-JUN-2003.
 PD 30-JUN-1998; 98US-00107532.
 PF 02-JUL-1997; 97US-0051571P.
 PR 14-MAY-1998; 98US-0085598P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 PI Doucette-Stamm LA, Bush D;
 DR WPI, 2003-799836/75.
 XX N-PSDB; ADC91461.
 PT New isolated nucleic acid derived from Enterococcus faecium encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.
 XX
 PS Example 1, SEQ ID NO 4742; 243pp; English.
 XX
 CC The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridizing to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acid is useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteremia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The

CC nucleic acid is useful for recombinant production of *Candida albicans* -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.
 XX
 SQ Sequence 378 AA;
 Alignment Scores:
 Pred. No.: 0.181 Length: 378
 Score: 86.00 Matches: 39
 Percent Similarity: 42.78 Conservative: 25
 Best Local Similarity: 26.08 Mismatches: 56
 Query Match: 9.24 Indels: 30
 DB: 7 Gaps: 8
 US-10-768-093-4 (1-527) x ADC95115 (1-378)
 QY 90 ATTCTTGCTTCATTCGGACGCCATCCGAAACAGAAATAGCGACTTAACTCCAGTA 149
 Db 158 VALLEUALAALAAAGLYALAGLYTHRGLNGLYserPTyrTHRGlyTHRSerPhe----- 175
 QY 150 TCAAGGACTATTTCAAAAAGTTTTCGACCTGACACGACGAAATACAGCCTTCTTGGT 209
 Db 176 --ThrsIleIleGLYHISGLYARGPHERTHRPROGLUALAGLYALGlnPROHAlIleGLY 194
 QY 210 GAAAATGTTGGAAG---GAAGSAGCTTTATTTAGTGTCGACTTAAGTCTGCTGAA 266
 Db 195 Lys---ValGLYHISThrGLNGLNValGlnGLNAlaLYSValGlnValIleLeuPROGLU 213
 QY 267 AATGTATCCAGGTACCGCTAC-----CCTGTT 296
 Db 214 ThrIleGLYHISGlnValIleGlnAlaMetARGSERAlaHISProTYRGLUInPROAla 233
 QY 297 TATGAT-----GAAGTTATGGTGTAGACGACTAATAAT 332
 Db 234 TYRAspLeuPHEAlaIleAspGLUProVALGlnMetPHEGLYLeuGLYARGLYAL----- 251
 QY 333 ACCGCTGATGCTTCCCATCATATATCTACACGATTTGTGATGAGAAAAGGAAAAATG 392
 Db 252 ---GLYGLUleuPROGLNGLUlePHEIleGLUALAAspValGLNValLYSGLUALa 270
 QY 393 TTAAGAATCATGTGTGACAGGTTACA-----CCTAATCAACAATACTTTTAA--- 443
 Db 271 PHEGlnLeuAspGLYLeuARGIleVALGlnPROLYSAsnAlaLYSserValLYSARG 290
 QY 444---GGGCTGAATTAATACAGCGGGAANA 470
 Db 291 IleaIleCysGLYserGLYGLNLYS 300
 RESULT 13
 ID AAB49208 standard; protein, 506 AA.
 XX AAB49208;
 AC
 DT 12-MAR-2001 (first entry)
 XX
 DE V. cholerae VPI phage conserved gene ALda protein.
 XX
 KM Bacteriophage; pathogenicity island; vaccine; allergy.
 XX
 OS Unidentified.
 FN WO200067784-A1.
 XX 16-NOV-2000.
 PD 10-MAY-2000; 2000WO-US012580.
 PF 10-MAY-1999; 98US-0133373P.
 XX
 PR 10-MAY-1999; 98US-0133373P.
 XX
 PA (UYMA-) UNIV MARYLAND BALTIMORE.

```

XX  Karoalle DKR;
PI  WPI/ 2001-122703/13.
DR  Isolated bacteriophage encoded by a pathogenicity island of a pathogenic
XX  bacterium, useful for generating bacteriophage-base vaccines or vectors,
XX  e.g. to treat allergies.
XX  Disclosure, Fig 3, 59pp; English.
XX  The present invention relates to an isolated bacteriophage encoded by a
XX  pathogenicity island of a pathogenic bacterium. The bacteriophage may be
XX  used in pharmaceuticals e.g. for generating an immune response,
XX  especially to produce antibodies, as a bacteriophage-based vaccine
XX  against pathogenic bacteria expressing virulence factors, as a vaccine
XX  vector, e.g. to prevent or treat allergies, diseases or other
XX  pathological conditions
XX  Sequence 506 AA;
SQ
Alignment Scores:
Pred. No.: 0.279 Length: 506
Score: 85.00 Matches: 51
Percent Similarity: 36.2% Conservative: 17
Best Local Similarity: 27.1% Mismatches: 62
Query Match: 9.1% Indels: 58
DB: 4 Gaps: 10
US-10-768-093-4 (1-527) x AAB49208 (1-506)
OY 78 GGTTAATTCATTTGCTTCATTCGCGACCCAGTCCAGAAACAGAAATAGGATTA 137
DB 211 GYVALLIIEAENVAlValaNglyPheGlySerGluAlaGlyAaMaLaLeuAlaThrSer 230
OY 138 AAC-----TTCCAGATCAACAGCATTTTCMAAAAGTTTTTT----- 176
DB 221 GlnAglIleAapIleuAlaApeThrGlySerThGluIleGlyAaHisIleuLys 250
OY 177 GCACCTGAACCAAGATACAGCCTTTGTTGTAAGAAATGTTGGAAG----- 224
DB 251 GYVAlAlaAaPAnleuIleProSerThrIleGluLeuGlyGlySerProAenile 270
OY 225 -----GAGAGAGCT 233
DB 271 TyrPheProAaPilePheSerHisGluAaPglntYrLeuAaPlyScyIleGluGlyAla 290
OY 234 TTATTATTAGTGTGAAC-----TTAAGTCTTCGTAAGAAATGATCCAGATA 281
DB 291 LeuLeuAlaPhePheAaNglyGlyValCysThrCysPro-----SerArgile 307
OY 282 ACCGCTTAC--CCTGTTATGAT-----GAGATTAATGGGTTA 317
DB 308 LeuValHisGluSerIleTyrgluYwpheIleAlaIleIleGluArgValAlaLeu 327
OY 318 GGAACAGTGTAAATACCGCTGCTTCCCAATCAATATCAACAGATTTGATGAG 377
DB 328 IleYvGlnGlyAaPProleuAaPThrGluThGlnIleGlyAlaGlnValSerIysGlu 347
OY 378 AAAGGAAAAAATGTTA-----AAAGATCATGGTGCAGAGGTAA 419
DB 348 GlnTyAaPlyIleuGlyTyrlleGlnIleGlyLysAaPglndlyAlaGluLeuile 367
OY 420 -----CCTAATCAACAAATTAATTTAAACCGCTGATTAATCAACGGGGGA 467
DB 368 PheGlyGlyHisPheAaNglyAaNgln-----GluAaPtyrLeuSerGlyGly 382
OY 468 AAAAAAATTCCTCGAATATG 491
DB 383 TyrTyrlleYwpheThrIleuPhe 390

```

```

ID  ABU49234 standard; protein; 541 AA.
XX  AC  ABU49234;
XX  AC  19-JUN-2003 (first entry)
XX  DR  Protein encoded by Prokaryotic essential gene #34761.
XX  DE  Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX  KW  Vibrio cholerae.
XX  OS  MO20027183-A2.
XX  PN  03-OCT-2002.
XX  PD  21-MAR-2002; 2002WO-US009107.
XX  PF  21-MAR-2001; 2001US-00815242.
XX  PR  06-SEP-2001; 2001US-00948993.
XX  PR  25-OCT-2001; 2001US-0342923P.
XX  PR  08-FEB-2002; 2002US-00072851.
XX  PR  06-MAR-2002; 2002US-0362699P.
XX  (ELIT-) ELITRA PHARM INC.
XX  Wang L, Zamudio C, Malone C, Hasselbeck R, Ohlsen KL, Zyskind JM;
XX  Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX  WPI; 2003-029926/02.
XX  N-PSDB; ACA53104.
XX  New antisense nucleic acids, useful for identifying proteins or screening
XX  for homologous nucleic acids required for cellular proliferation to
XX  isolate candidate molecules for rational drug discovery programs.
XX  Claim 25; SEQ ID NO 77158; 1766pp; English.
XX  The invention relates to an isolated nucleic acid comprising any one of
XX  the 623 antisense sequences given in the specification where expression
XX  of the nucleic acid inhibits proliferation of a cell. Also included are:
XX  (1) a vector comprising a promoter operably linked to the nucleic acid
XX  encoding a polypeptide whose expression is inhibited by the antisense
XX  nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX  polypeptide or its fragment whose expression is inhibited by the
XX  antisense nucleic acid; (4) an antibody capable of specifically binding
XX  the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX  proliferation or the activity of a gene in an operon required for
XX  proliferation; (7) identifying a compound that influences the activity of
XX  the gene product or that has an activity against a biological pathway; (8)
XX  required for proliferation, or that inhibits cellular proliferation; (9)
XX  identifying a gene required for cellular proliferation or the biological
XX  pathway in which a proliferation-regulated gene or its gene product lies
XX  or a gene on which the test compound that inhibits proliferation of an
XX  organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX  compound's activity; (11) a culture comprising strains in which the gene
XX  product is overexpressed or underexpressed; (12) determining the extent
XX  to which each of the strains is present in a culture or collection of
XX  strains; or (13) identifying the target of a compound that inhibits the
XX  proliferation of an organism. The antisense nucleic acids are useful for
XX  identifying proteins or screening for homologous nucleic acids required
XX  for cellular proliferation to isolate candidate molecules for rational
XX  drug discovery programs, or for screening homologous nucleic acids
XX  required for proliferation in cells other than S. aureus, S. typhimurium,
XX  K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX  the target prokaryotic essential genes. Note: The sequence data for this
XX  patent did not form part of the printed specification, but was obtained
XX  in electronic format directly from WIPO at
XX  ftp.wipo.int/pub/published_poc_sequences
XX  Sequence 541 AA;
SQ
Alignment Scores:

```

Pred. No.: 0.288 Length: 541
Score: 85.00 Matches: 51
Percent Similarity: 36.2% Conservative: 17
Best Local Similarity: 27.1% Mismatches: 62
Query Match: 9.1% Indels: 58
DB: 6 Gaps: 10

US-10-768-093-4 (1-527) x ABU49234 (1-541)

QY 78 GGTAAATTCCTTCTTCCTTCATTCGGGACCCATGCCAGACAGAAATAGCACTAAA 137
DB 246 G1YVAl1leAenVal1aenG1YpHeG1YserG1YAlaG1YAsnAlaLeuAlaThSer 265
QY 138 AAC-----TTCACAGTATCAACGACTATTTCAAAAAGTTTTTT--- 176
DB 266 G1NArg1leAer1YleuAla1erPhe1erTh1erG1YleuG1YAsnH1leu1Ys 285
QY 177 GCACCTGAACCAAGATACAGCCTTTCTTTGGTGAAAATGTTGGAAAG----- 224
DB 286 CysAla1a1aerAen1eul1erProserTh1leG1YleuG1Y1YserProAsn1le 305
QY 225 -----GAGGAGACT 233
DB 306 TyrPheProAer1lePheSerH1eG1YAsp1Y1YleuAer1YsCys1leG1YAla 325
QY 234 TTATTAATTTAGTGTAC-----TTAACTGTTCCGAAAATGTAATCCACGTA 281
DB 326 Leu1eul1a1erPheAen1eul1eG1YAla1YserTh1erCysPro-----SerTh1er1le 342
QY 282 ACGGCTTAC-----CCTGTTATGAT-----GAAATTAATGGTTA 317
DB 343 LeuVal1H1eG1Yser1le1Yer1YserPhe1leAla1Yle1leG1YArgVal1a1eul 362
QY 318 CGAGCAGTACGTAATACCGCTGATCTCCCAATCAATATACAGATGTTGATGAG 377
DB 363 1le1YserG1YAsp1YPro1eul1a1erTh1erG1Y1YleuG1YAlaG1YVal1eul1Yser1Ys1u 382
QY 378 AAAGGAAAAAATGTTA-----AAAGATCATGCTGACAGGTTACA 419
DB 383 G1N1YrAer1Yle1eul1eG1Y1Yr1leG1N1leG1Y1YsAerG1YAlaG1Yleu1e 402
QY 420 -----CCTAATCAACAAATTAATTTAAAGCGCTGAATTAATTAATGCGGGA 467
DB 403 PheG1Y1YH1e1YProAenAenG1N-----G1YAsn1Y1Yleu1YserG1Y1Y 417
QY 468 AAAAAAATATCTCCTCGAATATAT 491
DB 418 Tyr1Yr1le1Ys1YPro1Yr1leuPhe 425

RESULT 15

ID ADR86415 standard, protein; 1239 AA.

ADR86415;

AC ADR86415;

DT 04-NOV-2004 (first entry)

DE Aspergillus fumigatus essential gene protein #465.
KW Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
drug screening.

OS Aspergillus fumigatus.

PN WO2004067709-A2.

PD 12-AUG-2004.

PF 16-JAN-2004; 2004WO-US001099.

PR 17-JAN-2003; 2003US-0441281P.

PR 13-JUN-2003; 2003US-0478196P.

XX

PA (ELIT-) ELITRA PHARM INC.
PA (ELIT-) ELITRA CANADA LTD.
XX
XX Jiang B, Hu W, Lemieux S, Roemer T;
PI
XX WPI: 2004-5594200/57.
DR N-PSDB; ADR85828.

PT New purified or isolated Aspergillus fumigatus nucleic acid molecule
encoding a gene product, useful for diagnosing and/or treating invasive
fungal infections, such as Farmer's lung disease.

PS Claim 1; SEQ ID NO 3465; 164pp; English.

XX The present invention relates to Aspergillus fumigatus genes that are
XX essential and are potential targets for drug screening. The methods and
XX compositions of the present invention are useful for diagnosing and/or
XX treating invasive Aspergillus fumigatus infection, including the allergic
XX forms of the disease, such as Farmer's lung disease. They can also be
XX used in various drug discovery purposes, such as expression of the
XX recombinant protein, hybridization assay and construction of nucleic acid
XX arrays. The present sequence represents an Aspergillus fumigatus
XX essential gene protein sequence, used during diagnosis and drug
XX development in the invention. These genes share a high degree of sequence
XX conservation with known essential genes of candida albicans. The sequence
XX data for this patent is not represented in the printed specification, but
XX was obtained in electronic format from WIPO.

SQ Sequence 1239 AA;

Alignment Scores:

Pred. No.: 0.488 Length: 1239
Score: 84.50 Matches: 35
Percent Similarity: 44.0% Conservative: 20
Best Local Similarity: 28.0% Mismatches: 53
Query Match: 9.0% Indels: 17
DB: 8 Gaps: 6

US-10-768-093-4 (1-527) x ADR86415 (1-1239)

QY 150 TCAACGACTATTTCGAAAAGTTTTTTCACACT-----GAA 185
DB 1031 AlArThr1H1e1Val1aenA1Yser1YserPro1a1a1eG1YAlaenG1YTh1YsAla 1050
QY 186 CCACGATACAGCCTTTCTTTGGTGAATAATGTTGAAGAAGAGCTTATTTATTTAGT 245
DB 1051 ProserValG1YPro-----G1YG1YTh1a1Ys1YG1YAsn1Yser1YAsn1YArgTh1Ydeu 1067
QY 246 GTGAATTAATCTGTTCTGAAAATGTAATCCAGTAACGGCTACCCGTTATGATGAA 305
DB 1068 G1Yleu1eul1aenVal1YProAerTh1YVal1aenA1Yser1Yser1YAsn1YArgTh1Ydeu 1086
QY 306 GATTAATGGG---TTAGACGAGTACGTAATAATACCGCTGATGCTTCCCAATCAATATATC--- 359
DB 1087 Pro1Yr1eG1YPro1eul1e1Yle1Yle1Yle1YleuA1Yr1YProAerH1e1YserG1YAla1le1YValG1Y 1106
QY 360 TACCAATGTTGTATGAGAAAGGAAAAAATGTTAAAGATCATGCTGACAGGTTAC 419
DB 1107 PheAlaAerValAenH1eAlaG1Y1Yser1Yser1YleuG1Y1YleuG1Y1YleuG1Y1Yleu 1126
QY 420 CCTAATCAACAAAT-----ACTTTAAAGCGCTGAATTAATTAATGCGGGA 470
DB 1127 ProG1Y1Yr1YleuH1e1YserG1YTh1YVal1e1YserG1Yleu1YserG1Y1Yleu1Ys 1146
QY 471 AAAATATCTCCTGGA 485
DB 1147 1YserG1Y1YProG1Y 1151

Search completed: April 10, 2006, 10:00:10
Job time : 140.5 secs

T09437
Probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) alda [similarity] - Vibrio cholerae
C/Species: Vibrio cholerae
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C/Accession: T09437
R/Karaozler, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kaper, J.B.; Reeves, P.
Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998
A/Title: A Vibrio cholerae pathogenicity island associated with epidemic and pandemic st
A/Reference number: 216572; MUID:98169509; PMID:9501228
A/Accession: T09437
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-506 <KAP>
A/Cross-references: UNIPROT:P23240; UNIPARC:UP10000000A06; EMBL:AF034434; NID:G3004923;
A/Experimental source: strain N16961
C/Genetics:
A/Gene: alda
A/Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic st
A/Function:
A/Description: catalyzes the oxidation of aldehydes
C/Superfamily: NAD-dependent aldehyde dehydrogenase; aldehyde dehydrogenase homology
C/Keywords: NAD; oxidoreductase

Alignment Scores:
Pred. No.: 1.45 Length: 506
Score: 85.00 Matches: 51
Percent Similarity: 36.2% Conservative: 17
Best Local Similarity: 27.1% Mismatches: 62
Query Match: 9.1% Indels: 58
DB: 2 Gaps: 10

US-10-768-093-4 (1-527) x T09437 (1-506)

OY 78 GGTAAATTCGAATCTGCTTCATTCGGAGCCATGCCAGACAGAAATAGCGACTAA 137
DB 211 GYVALLIENAVENVAIENGLYPHEGLYSERGIWAAGLYASNALEUVALHISER 230
OY 138 AAC-----TTCCAGATCAAGACTATTTCAAAAAGTTTTTT-- 176
DB 231 GlnrgliearplyleuvalaRhetrglyserthrguileglyAsnhi8tleuuls 250
OY 177 GCACCTGAACAGAAATACAGCTCTTTTGGTGAATGTTGGAAG----- 224
DB 251 CyslaalaRpaenleuileproserthrliegileuglylyserprosnile 270
OY 225 -----GAAGAGCT 233
DB 271 TyrpheproapRlepheserhi8gluabrgintyRleuabplyscysilegileuylala 290
OY 234 TTATTAATTAAGTGTGAC-----TTAACTGTTCTGAAATGTAATCCAGGTA 281
DB 291 leuenuaRphesRenglnglylualcysethrcyspro-----Serhrgile 307
OY 282 ACGGCTAC--CCTGTTATGAT-----GAAGATTATGGGTTA 317
DB 308 leuvalhi8gluseriletyrglyuRphelalealysileilegluargvalaleu 327
OY 318 GGAGCAGTAGTAATACCGCTGATGCTCCCAATCAATATACCATGTTGATGAG 377
DB 328 ileuglnglyAsnproleuabRphthrgluhrginileglyalaglnvalserlyseu 347
OY 378 AAAGGAAAAAATGTTA-----AAAGATCATGTCAGAGGTTACA 419
DB 348 GlnTyRabplysileleuglytyrileglnileglylysaRpglnglyalagileuile 367
OY 420 -----CCTAATCAACAATACTTTAAAGCGCTGAATTATACAGCGGGAA 467
DB 368 pheglylyhi8proabna8ngln-----GluabntyRleu8erlygly 382
OY 468 AAAAAATATCTCCTGGAATATAT 491
DB 383 TyrtYrillely8prothRleuphe 390

RESUR 5

A82276

aldehyde dehydrogenase VC0819 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Mar-2004

C/Accession: A82276

R/Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: A82276

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-541 <HEI>

A/Cross-references: UNIPARC:UP10000164B65; GB:AE004167; GB:AE003852; NID:99655268; PIDN:

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C/Genetics:
A/Gene: VC0819
A/Map position: 1
C/Superfamily: NAD-dependent aldehyde dehydrogenase; aldehyde dehydrogenase homology

US-10-768-093-4 (1-527) x A82276 (1-541)

OY 78 GGTAAATTCGAATCTGCTTCATTCGGAGCCATGCCAGACAGAAATAGCGACTAA 137
DB 246 GYVALLIENAVENVAIENGLYPHEGLYSERGIWAAGLYASNALEUVALHISER 265
OY 138 AAC-----TTCCAGATCAAGACTATTTCAAAAAGTTTTTT-- 176
DB 266 GlnrgliearplyleuvalaRhetrglyserthrguileglyAsnhi8tleuuls 285
OY 177 GCACCTGAACAGAAATACAGCTCTTTTGGTGAATGTTGGAAG----- 224
DB 286 CyslaalaRpaenleuileproserthrliegileuglylyserprosnile 305
OY 225 -----GAAGAGCT 233
DB 306 TyrpheproapRlepheserhi8gluabrgintyRleuabplyscysilegileuylala 325
OY 234 TTATTAATTAAGTGTGAC-----TTAACTGTTCTGAAATGTAATCCAGGTA 281
DB 326 leuenuaRphesRenglnglylualcysethrcyspro-----Serhrgile 342
OY 282 ACGGCTAC--CCTGTTATGAT-----GAAGATTATGGGTTA 317
DB 343 leuvalhi8gluseriletyrglyuRphelalealysileilegluargvalaleu 362
OY 318 GGAGCAGTAGTAATACCGCTGATGCTCCCAATCAATATACCATGTTGATGAG 377
DB 363 ileuglnglyAsnproleuabRphthrgluhrginileglyalaglnvalserlyseu 382
OY 378 AAAGGAAAAAATGTTA-----AAAGATCATGTCAGAGGTTACA 419
DB 383 GlnTyRabplysileleuglytyrileglnileglylysaRpglnglyalagileuile 402
OY 420 -----CCTAATCAACAATACTTTAAAGCGCTGAATTATACAGCGGGAA 467
DB 403 pheglylyhi8proabna8ngln-----GluabntyRleu8erlygly 417
OY 468 AAAAAATATCTCCTGGAATATAT 491
DB 418 TyrtYrillely8prothRleuphe 425

A/Molecule type: DNA
 A/Residues: 1-740 <WE11>
 A/Cross-references: UNIPROT:Q46266; UNIPARC:UPI0000131758; EMBL:X93463; NID:gl072360; PI
 A/Accession: PC6004
 A/Molecule type: protein
 A/Residues: 1-740 <WE12>
 A/Cross-references: UNIPARC:UPI0000131758
 C/Comment: This enzyme must be activated by the lyase-specific pyruvate formate-lyase-ac
 ite and causes peptide cleavage.
 C/Genetics:
 A/Gene: pfl
 A/Start codon: TTG
 C/Function:
 C/Complex: homodimer
 A/Description: catalyzes the reversible conversion of acetyl-CoA and formate into pyruv
 A/Pathway: anaerobic glucose metabolism
 C/Superfamily: formate-C-acetyltransferase 1; glycyl radical homology
 C/Keywords: acyltransferase; coenzyme A; homodimer; lyase; thioester bond
 F/681-740/Domain: glycyl radical homology <GPR>
 F/405/Active site: Cys (cysteine thiol radical intermediate) #status predicted
 F/406/Active site: Cys (S-acetylcysteine intermediate) #status predicted
 F/715/Active site: Gly (stable glycyl radical) #status predicted

Alignment Scores:
 Pred. No.: 1.98 Length: 740
 Score: 84.00 Matches: 39
 Percent Similarity: 38.64 Conservative: 25
 Best Local Similarity: 23.54 Mismatches: 68
 Query Match: 8.94 Indels: 34
 DB: 1 Gaps: 7

US-10-768-093-4 (1-527) x JC6010 (1-740)
 Oy 93 CTTCCTTCATTCGGAGCGCATCCGACAGAAATAGGACTAAAACCTCCAGTATCA 152
 Db 326 ILEAEGLYVALGTYLLEAPSPGLYARSGERLEUVALTHRLYSANSEPHERGTYRLEU 345
 Oy 153 ACGACTATTTCAGAAAGTTTTCACCTGACCAACAGCAATACAGCTCTTTGGTGA 212
 Db 346 HIRHREULEANLEUGLYSERLAPROGLUPROAMMETHVALLEUTRISERGLU 365
 Oy 213 AATGTGGAAGAAGAGCTTATTATTAGTGGAACCTTACTGTTCTGAAATGTA 272
 Db 366 ANLEU---PROGLUSERPHELSELYSEPCYALAGLUMESERILELEUTHRASPGR 384
 Oy 273 TCCAGAGTA-----ACGCTACCCCTGTTATGATGAGATTAGGGTTA--- 317
 Db 385 ILEGINTYRGLUANAPSPAPRILEMETARSPROILETYRGLYASPSPTALALEALA 404
 Oy 318 -----GGACGACTAGTAAT 332
 Db 405 CYSCYVALISERALEMETARGLYVALYSAPMETGLINPHEPCGLYALARGCYASHN 424
 Oy 333 ACCGCTGATGCTTCCCAATCAATAATCTACAGATTGTTGATGAGAAAAGGAAAAAATG 392
 Db 425 LEUVALYSYCYLEULEULEUALALEANGLYVALASPGILYS----- 440
 Oy 393 TTTAAAGATCATGTGCGACAGGTTACACCTAAT---CAACAATAACTTTTAAGCGCTG 449
 Db 441 -----LYSGLYTLELYVALAIPROASPILEGULPROILETHRASPGLUVALLEU 457
 Oy 450 AATTACTACGGGGAAGAAAAAATATCTCTCGAATATATACGAT---CAGTTATG 506
 Db 458 ASPTYR-----GLULYSVALYSGLUANTYRPHELYVALLEUGLUTYRNET 473
 Oy 507 GTTGTTACTATGTATAC 524
 Db 474 ALAELYLEUTYRVALASN 479

RESULT 9
 JC7604
 CD86 spliced variant CD86 deltaCTM isoform - human
 C/Species: Homo sapiens (man)

C/Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 18-Nov-2002
 C/Accession: JC7604
 R/Magistralli, G.; Caron, G.; Gauchat, J.F.; Jeannin, P.; Bonnefoy, J.Y.; Delneste, Y.
 Biochem. Biophys. Res. Commun. 280, 1211-1215, 2001
 A/Title: Identification of an alternatively spliced variant of human CD86 mRNA.
 A/Reference number: JC7604; MUID:21092744; PMID:11162656
 A/Accession: JC7604
 A/Molecule type: mRNA
 A/Residues: 1-275 <MAG>
 A/Cross-references: UNIPARC:UPI0000178C43
 C/Comment: This CD86 variant expressed by activated human monocytes, is a costimulatory
 C/Genetics:
 A/Gene: cd8deltaCTM
 C/Superfamily: B-lymphocyte restricted antigen B7
 C/Keywords: immune response

Alignment Scores:
 Pred. No.: 1.86 Length: 275
 Score: 83.50 Matches: 39
 Percent Similarity: 39.8% Conservative: 29
 Best Local Similarity: 22.8% Mismatches: 58
 Query Match: 8.94 Indels: 45
 DB: 2 Gaps: 6

US-10-768-093-4 (1-527) x JC7604 (1-275)
 Oy 66 AAGAAACAATTGGTTTAACTTAAT-----CTTGCT 98
 Db 109 LYSLSYSPROTHRGLYMETILEARGILEHISGLINMETANSESRGILEUSERVALLEUULA 128
 Oy 99 TCATTGGGAGCCATGCGACAGCAAGAAATAGCG-----ACTAAAACTTCCAGTATCA 152
 Db 129 ASNPHESESRGLNPROGLULLEVALPROILESESRANLETHRGLUASVALTYRILEASN 148
 Oy 153 ACGACTATTTCAGAAAGTTTTCACCTGACCAACAGCA----- 191
 Db 149 LEUTHRYSSESRILEHLEGLYTRPROGLUPROLYLSMETSERVALLEULEUARG 168
 Oy 192 -----ATACAGCTTCTTTGTTGTAAGTGTGA 221
 Db 169 THRYSANSESRTHRILEGLUTYRASPGLYILEMETGLINYSERGLNASPANVALTHR 188
 Oy 222 AAGAGAGCTTATTATTATTAGTGGAACCTTACTGTAAGATGTATCCAGGTA 281
 Db 189 GLUENTYRASPVALISERILESESRLEUSERVALSERPHEPROASPVALTHRSESRMET 208
 Oy 282 ACGCTACCCCTGTTATGATGAGAT-----TATGGGTTA 317
 Db 209 THRILEPCYILELEUGLUTHRASPGLYTHRARGLEULEUSERSESRPROPHESERTILE 228
 Oy 318 GGACGACTAGTAATACCGCTGATGCTTCCCAATCAATAATCTACAGATTGTTGATGAG 377
 Db 229 GLY-----THRANTRHMERGLUARGLUGLUBER-----GLU 239
 Oy 378 AAGAGGAAAAAATGTAAAAGATCATGTGTCAGAGGTTACACTAATCAACAATAACT 437
 Db 240 GLNTHRYLSYASARGLYLLEHISLERPROGLUNYRGSERASPGILUALAGLARGVAL 259
 Oy 438 TTTAAGCGCTGAATTATCTACGCGGGGAANA 470
 Db 260 PHEUSSESRLEYSRTHRSESRCYASHPYS 270

RESULT 10
 AF1226
 hypochemical protein lmo1214 [imported] - listeria monocytogenes (strain EGD-e)
 C/Species: listeria monocytogenes
 C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C/Accession: AF1226
 R/Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fathi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

Ok, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1226
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-183 <GLA>
 A:Cross-references: UNIPROT:Q8YQ08; UNIPARC:UPI000055539; GB:NC_003210; PIDN:CAC99292.1
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo1214

Alignment Scores:
 Pred. No.: 1.95 Length: 183
 Score: 83.00 Matches: 38
 Percent Similarity: 37.9% Conservative: 26
 Best Local Similarity: 22.5% Mismatches: 53
 Query Match: 8.8% Indels: 52
 DB: 2 Gaps: 7

US-10-768-093-4 (1-527) x AF1226 (1-183)

QY 399 CTTTAAACATTTTTCCTTCATCAACAATCGTAGATTAATGATGGGAAG--- 343
 DB 5 ValValIysPhePheThValAspMetGluValGluGluAlaTyrLeuAsnGluMet 24
 QY 342 -----CATCAGCGGTAATTTACTAGTCGCTCAACCCATATCTTCAT 301
 DB 25 AlaGluAsnGlyTyrPhePheGlnLysTyr-----LysSerPheLysTyrHis 40
 QY 300 CATAAACGGGTAGACCGTTACT-----GGGATACATTTTCAGAAACAGTTAACT 250
 DB 41 PheGluGlnGlyValProAlaLysTyrSerTyrAlaIleAspPheLysGluAsnGluGly 60
 QY 249 TCACACTAAATATATAAGCTCTT----- 226
 DB 61 AspGluGluAlaTyrLysThrLeuLeuGluAspAlaGlyTyrGluThrValPheSerTyr 80
 QY 225 -----CCTTTCACACATTTTCACCA-----AAG 202
 DB 81 ProValLeuAsnGlyAsnTyrPheArgLysAlaValAlaProGlyGluThrGlu 100
 QY 201 AAGCGTGAATC-----GTGCTCAGCGTCAAAAACCTTTGAAATAG 157
 DB 101 GluAlaIlePheThAspGluThrSerLeuIleGlnLeuTyrLysAsnIleArgLysArg 120
 QY 156 TCGTGATACCTGGAAAGTTTATGTCGATTTCTGTCGATGCGTGCAGATGAAG 97
 DB 121 TrpThrIlePheGlyAlaIleValSerLeuPhePhe----- 133
 QY 96 CAAGAATTAGATTAAACCAATGTTTCTCATATAACACATTAATATGCTAATATC 37
 DB 134 -----ValGluLeuLeuIleValPheGlnIleGluAsnHisIleGlyLeuAlaThrPhe 151
 QY 36 ATTTTATTTGTATACATATTCGA 10
 DB 152 IlePheIleIlePhePheIleLeuVal 160

RESULT 11
 T16580
 Hypothetical protein K07E12.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T16580
 R:Fulton, L.
 submitted to the EMBL Data Library, May 1994
 A:Description: The sequence of C. elegans cosmid K07E12.
 A:Reference number: Z18540
 A:Accession: T16580
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-13055 <FULL>
 A:Cross-references: UNIPROT:Q09165; UNIPARC:UPI000017CF3A; EMBL:U00054; NID:G485140; PID

A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:K07E12.1
 A:Introns: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3; ; 6014/3; 6159/3; 6665/2; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1225

Alignment Scores:
 Pred. No.: 4.26 Length: 13055
 Score: 83.00 Matches: 34
 Percent Similarity: 42.2% Conservative: 23
 Best Local Similarity: 25.2% Mismatches: 50
 Query Match: 8.8% Indels: 28
 DB: 2 Gaps: 6

US-10-768-093-4 (1-527) x T16580 (1-13055)

QY 144 CCAGTATCAACGACTATTTCAAAAAGTTTTCACCTGACCAAGATA-----CAG 197
 DB 10401 ProLeuGlyThrAspSerSerGlySerTyrIleThrGluAspGlyGlnLeuValGlyLys 10420
 QY 198 CTTCTTTTGGTAAATGTTTGGAAAGAGAGACTTATTTATTTAGTGAACCTAACT 257
 DB 10421 AspGluGluGlyLysProValGlyProAspGlyGlnValLeu----- 10434
 QY 258 GTTCTGAAATGTATGCCAGTAGAACGCTCACCTGTTATGTAAGATTATGCGTTA 317
 DB 10435 ---ProThrAspSerAlaGlyHisTyrValTyrProIle-----ThrGlyAla 10449
 QY 318 GCAGCACTAGTAATTAATCCGCTGATGCTTCCCAATCAATATCTACAGATTGTGATGAG 377
 DB 10450 AspArgGlnIleLeuThrThrAspAlaIleGlyLysProIleTyrSerValPheAsnGlu 10469
 QY 378 AAGGGAAAAAATGTTAAAGAT-----CATGCGCAGAGTT 416
 DB 10470 AspGlyIleGlnLeuProThrAspSerSerGlyTyrAlaIleGlyHisAspGlyGlnLeu 10489
 QY 417 ACACCTATCAACAATAATCTTTAAAGCGCTGATTAATCAAGCGGGAATAAATA 476
 DB 10490 ValProThrGluSerThrAsnGlyValProLeuAsn-----LysAspGly 10504
 QY 477 TCTCTGGAAATATATAACGATCAGATTATGCTGTTACTATGTA 521
 DB 10505 ThrProLeuProThrAsnAspSer-----GlyHisPheVal 10516

RESULT 12
 G64463
 Hypothetical protein MJ1312 - *Methanococcus jannaschii*
 C:Species: *Methanococcus jannaschii*
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
 C:Accession: G64463
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurest, W.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: G64463
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-321 <BLU>
 A:Cross-references: UNIPROT:Q58708; UNIPARC:UPI000013ATC6; GB:U67571; GB:L77117; NID:G15
 C:Genetics:
 A:Map position: FOR1257323-1258288

Alignment Scores:
 Pred. No.: 3.51 Length: 321
 Score: 81.00 Matches: 46
 Percent Similarity: 44.6% Conservative: 29
 Best Local Similarity: 27.4% Mismatches: 56
 Query Match: 8.6% Indels: 37
 DB: 2 Gaps: 10

QY 180 CCT---GAACACGAAATACAGCTTCT
 |||||
 Db 204 ProValAlaIProLysIleProProAlaIProAlaIAlaLeuPheAlaGlyAla 223
 QY 207 GGTGAAAATGTTGAAAGGAAAGGAGCTTTATTTATTTAGTGTAACTTAACCTGTCCTGAA 266
 |||||
 Db 224 TrpGluAenTyrgLYAlaGluGlnPheLeuAenTyrgLY-----ArgLeuProGly 240
 QY 267 AATGATATCCAGAGTAACGCTTACCTGTTTATGATGAAGAAATTAT-----GGGTTAGCA 320
 |||||
 Db 241 AsnLeuPheMetIleAsn--TrpProIleCysGlyAsnAspTyrgLYGluGlyValGly 259
 QY 321 CGACTAGTAATACCGCTGATGCT-----TCC 347
 |||||
 Db 260 ArgLeuIleGluSerAlaIleAlaIleAlaLysGlyAspPhePheGlnGluSerArgTrpHisser 279
 QY 348 CAATCAATTAATCTACACAGTTGTTGATGAGAAAGGAAAAAATGTTAAAAAGATCATGCT 407
 |||||
 Db 280 GlnAsnPheAlaHisPheIleGlnThrGlnLeuGlyAArg-----TyrgLY 295
 QY 408 -----GGAGGGTTAACACCTTAATCAACAAATTAACCTTTTAAGCGCTGAATAT----- 455
 |||||
 Db 296 LeuAlaGluGlnValPheProHisAlaSerSerAlaPheAlaLeuHisProTyrgLYAArg 315
 QY 456 -----ACTAGCGGGGAAAAAATAATATCTCT----- 482
 |||||
 Db 316 GluSerArgArgLeuValGlyLeuThrThrValArgGluGlnIleAspIleLeuProValPro 335
 QY 483 -----GGAATATATAACGATCAGGTTATGGTTGGTACTATGTAAAC 524
 |||||
 Db 336 GlyGluGlnValAlaSerLeuPheProAlaIleAlaIleGlyAenTyrgLYAlaAsn 354

RESULT 14
 C65088
 hypothetical protein b3013 - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
 C/Accession: C65088
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.: Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A./Title: The complete genome sequence of Escherichia coli K-12.
 A./Reference number: A64720; PMID:9278503
 A./Accession: C65088
 A./Status: preliminary; nucleic acid sequence not shown; translation not shown
 A./Molecule type: DNA
 A./Residues: 1-309 <BLAT>
 A./Cross-references: UNIPROT:Q46858; UNIPARC:UPI000013BF0E; GB:AE000383; GB:U00096; NID:9
 A./Experimental source: strain K-12, substrain MG1655
 C./Superfamily: Escherichia coli hypothetical protein b3013

Alignment Scores:
 Fred. No.: 3.93 Length: 309
 Score: 80.50 Matches: 41
 Percent Similarity: 43.2% Conservative: 32
 Best Local Similarity: 24.3% Mismatches: 57
 Query Match: 8.6% Indels: 39
 DB: 2 Gaps: 10

US-10-768-093-4 (1-527) x C65088 (1-309)

QY 63 ATGAGAAACAAATGGTTTAATTTCTTAATTCCTTGCTTCATTCGGCAGCAGCATGCAGAACCA 122
 |||||
 Db 1 MetuysIleIleLeuLeuPheLeuPheLeuIleAlaIleAlaLeuAlaSerPheThrValHisAlaGln-- 19
 QY 123 GAAATAGCAGCTAAAAAATCTCCAGATATCAACGACTATTTCCAAAAGTTT----- 173
 |||||
 Db 20 -----ProProSerGlnThrValGluGlnThrValArgHisIle 32
 QY 174 -----TTTGACACTGAACCAAGATATACGCTCTTTGTTGGTGAATAATGTTGAAG--- 224
 |||||
 Db 33 TyrGlnAenTyrgLYSerSerAlaThrAlaProTyrgPheGlyGluIleGlnGlyAlaArgAla 52

Search completed: April 10, 2006, 10:07:31
Job time : 46 secs

```
QY 225 -----GAGAGCTTATTATTAGTGTGACTTAAGTCTGTA 266
DB 53 lIethrserAlArGIleGInGInAlLeuThrLeuAlAspAsnLeuThrLeuProGly 72
QY 267 AATGTATCCAGTAAAGCTTACCGTGTATGAT---GAGATTATGCG----- 314
DB 73 AsnIIeGlyTrpLeuAspTyrAspProValCysAspCysGlnAspGlyAspLeuVal 92
QY 315 -----TTAGACGACTAGTAATACCGGTGATGCTTCCCAATCAATAATC---TAC 362
DB 93 LeuGluSerValAlaIleThrGlnThrAspAlaAspHisAlaAspAlaValAlaArgPhe 112
QY 363 CAGATTGTT---GATGAGAAAGGAAA-----AAAATGTTAAAGATCAT 404
DB 113 ArgIlePheIleAspAspGlySerGlnThrGlnThrLeuIleuIleValAlaGluAsn 132
QY 405 GGTGCA-----GAGTTACAGTAAATCAACAATAACTTTAAAGCGTGAAT 452
DB 133 GlyArgTrpValIleAspAspIleValSerAspHisGlySerValLeuGlnAlaValAsn 152
QY 453 TATACTACCGCGGAAAAAATAATCT 479
DB 153 ---SerGluAsnGluLeuThrLeuAla 160
```

RESULT 15

```
T47324
Hypothetical protein T12K4.110 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #ext_change 09-Jul-2004
C:Accession: T47324
R:Montfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Rudd, S.; Lemcke, K.; May
Submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24460
A:Accession: T47324
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-951 <MON>
A:Cross-References: UNIPROT:Q9M298; UNIPARC:UPI000009FF20; EMBL:AL138640
A:Experimental source: cultivar Columbia; BAC clone T12K4
C:Genetics:
A:Map position: 3
A:Intron: 93/1; 468/3; 527/3; 565/3; 600/3; 645/3; 664/2; 696/3; 735/3; 778/1; 821/3
A>Note: T12K4.110
```

Alignment Scores:

Pred. No.:	5.44	Length:	951
Score:	80.00	Matches:	25
Percent Similarity:	38.4%	Conservative:	8
Best Local Similarity:	29.1%	Mismatches:	29
Query Match:	8.5%	Indels:	24
DB:	2	Gaps:	3

US-10-768-093-4 (1-527) x T47324 (1-951)

```
QY 295 ACAGGAGTACCGCTTACCTGATACATTTTACAGAACGTTAAGTTCACACTTAATAT 236
DB 167 ThrGlyThrValLeuCyTrpGluLeuGlnAsnGlyValValSerPheThrLeuIleGly 186
QY 235 AAAGCTCCCT-----TCCTTCCAAATTTTACCAACCAAA 203
DB 187 ValAlaProAspThrGlyPheAsnThrSerIleValAsnIleProArgTrpSerProAsp 206
QY 202 GAAGCTGTATTCTGTGTTCAAGT----- 179
DB 207 GlyArgThrLeuAlaValProGlyLeuArgAsnAspValValMetCysAspArgPheThr 226
QY 178 GCAAAAAAAGCTTTTAAATAGTCGTTGATACATCGGAAGTTTTCGTTATTTCTGTT 119
DB 227 GlyGluLeuPheLeu-----ArgGlyAspHisLeuGlnAlaIleCysTyr 243
QY 118 CTGGCATGGCTGCCGAAT 101
DB 244 LeuThrTrpAlaProAsn 249
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 10, 2006, 09:55:45 ; Search time 35.5 seconds
(without alignments)
2094.723 Million cell updates/sec

Title: US-10-768-093-4
Perfect score: 939
Sequence: 1 TTGCACATTCAGATGTTA.....TTGTTACTAGTAACTAA 527

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 216643 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-DB=ABSB/ABSB/EPCC/US1078093/runac_10042006_090303_4519/app_query.fasta_1
-DB=UniProt -QMT=FASTA -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p
-USER=us1078093 @GSI 1.1.466 @runac 10042006 090303 4519 -NCPU=6 -ICPU=3
-NO MMAR -NEG SCORES=0 -NMT -DSRBLCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	83.6	154	1 F6A2_ECOLI	P53509 escherichia
2	741	78.9	154	1 F6A1_ECOLI	P53508 escherichia
3	126	13.4	167	1 F6B1_ECOLI	P53510 escherichia
4	118	12.6	167	1 F6B2_ECOLI	P53511 escherichia
5	92	9.8	440	2 Q54US8_DICDI	Q54US8 dictyostell
6	91	9.7	374	2 Q5XSO1_LEGPN	Q5XSO1 legionella
7	90.5	9.6	318	2 Q4K1L8_STRPN	Q4K1L8 streptococc
8	88.5	9.5	318	2 Q4K1O5_STRPN	Q4K1O5 streptococc
9	88.5	9.4	1007	2 Q37318_HABPV	Q37318 heliothis a
10	88	9.4	2542	2 Q54ED6_DICDI	Q54ED6 dictyostell
11	87	9.3	918	2 Q72P30_LEPIC	Q72P30 leptospira
12	87	9.3	918	2 Q8F7D7_LEPIN	Q8F7D7 leptospira
13	86	9.2	2552	2 Q54ED7_DICDI	Q54ED7 dictyostell
14	85.5	9.1	573	2 Q81ASO_PLAF7	Q81ASO plasmodium
15	85.5	9.1	718	2 Q9M2C6_ARATH	Q9M2C6 arabidopsis
16	85	9.1	506	1 ALDH_VIBCH	P23240 vibrio chol

17	85	9.1	506	2 Q7BGD6_VIBCH	Q7BGD6 vibrio chol
18	84.5	9.0	437	2 Q8MVR9_CLOSI	Q8MVR9 clostrich
19	84	8.9	356	2 Q5X787_LEGPA	Q5X787 legionella
20	84	8.9	439	1 ENO11_SCHPO	P40370 schizosacch
21	84	8.9	456	2 Q4S970_TETNG	Q4S970 tetradion n
22	84	8.9	740	1 PFL_CLOPA	Q4S266 clostridium
23	83.5	8.9	205	2 Q8ZST2_PYRAE	Q8ZST2 pyrobaculum
24	83.5	8.9	454	2 Q897U6_CLOTE	Q897U6 clostridium
25	83	8.8	183	2 Q72OL5_LISME	Q72OL5 listeria mo
26	83	8.8	183	2 Q8Y7O8_LISMO	Q8Y7O8 listeria mo
27	83	8.8	374	2 Q5WYN8_LEGPI	Q5WYN8 legionella
28	83	8.8	749	2 Q81IA3_BACCR	Q81IA3 bacillus ce
29	83	8.8	12268	2 Q8M0U8_CAEEL	Q8M0U8 caenorhabdi
30	83	8.8	13100	2 Q09165_CAEEL	Q09165 caenorhabdi
31	82.5	8.8	275	2 Q9BDN9_PAPAN	Q9BDN9 papio anubi
32	82.5	8.8	385	2 Q7RKX4_PAPAY	Q7RKX4 plasmodium
33	82.5	8.8	440	2 Q8A503_BACTN	Q8A503 bacteroides
34	82.5	8.8	499	2 Q8G8H8_PSEPK	Q8G8H8 pseudomonas
35	82.5	8.8	598	2 Q8A637_PROAC	Q8A637 proteobac
36	82	8.7	248	2 Q6H0E5_BACAN	Q6H0E5 bacillus an
37	82	8.7	248	2 Q63OP3_BACIZ	Q63OP3 bacillus an
38	82	8.7	248	2 Q6HAS8_BACHK	Q6HAS8 bacillus th
39	82	8.7	264	2 Q81JY4_BACAN	Q81JY4 bacillus an
40	82	8.7	268	2 Q4MG91_BACCE	Q4MG91 bacillus ce
41	82	8.7	279	2 Q84214_SBACT	Q84214 uncultured
42	82	8.7	287	2 Q72X92_BACCI	Q72X92 bacillus ce
43	82	8.7	420	2 Q68ET9_XENLA	Q68ET9 xenopus lae
44	82	8.7	524	2 Q6CDU2_YARLI	Q6CDU2 yarrowia li
45	82	8.7	610	2 Q5WV11_LEGPI	Q5WV11 legionella

ALIGNMENTS

RESULT 1
FEA2_ECOLI STANDARD: PRT: 154 AA.
ID F6A2_ECOLI
AC P53509;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE CS6 fibribial subunit A precursor (CS6 pilin).
GN Name=cs6a;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=E8775;
RA Wolf M.K., de Haan L.A.M., Casseels F.C., Willems G.A., Geestel E.C.M.,
RA Gaestera W., Warren R., Boedeker B.C.;
RU Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Fibribriae (also called pili), polar filaments radiating
CC from the surface of the bacterium to a length of 0.5-1.5
CC micrometers and numbering 100-300 per cell, enable bacteria to
CC colonize the epithelium of specific host organs.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
DR EMBL: U04846; AAB51361.1; -; Unassigned_DNA.
KW SIGNAL
FT SIGNAL 1 18 Potential.
FT CHAIN 19 154 CS6 fibribial subunit A.
SQ SEQUENCE 154 AA; 16940 MW; 421E223D9FASFCB8 CRC64;
Alignment Scores:
Pred. No.: 1,996-66 Length: 154
Score: 785.00 Matches: 154
Percent Similarity: 100.0% Conservative: 0

FT SIGNAL 1 21 Potential:
FT CHAIN 22 167 CS6 fimbrial subunit B.
SQ SEQUENCE 167 AA; 18022 MW; F9FFIB91C41848E CRC64;

Alignment Scores:

Pred. No.: 0.0032 Length: 167
Score: 126.00 Matches: 49
Percent Similarity: 45.8% Conservative: 28
Best Local Similarity: 29.2% Mismatches: 75
Query Match: 13.4% Indels: 16
DB: 1 Gaps: 8

US-10-768-093-4 (1-527) x FB82_ECOLI (1-167)

```

OY 63 ATGAGAAACAATAT---GGTTAATCTAATCTTCTGTCATTCGAGCCATCCAGCA 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 LeuylsylvilletteProalaileValleuilealaglythserGlyValValasmla 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 120 ACAGAAATAGGACTAAAACTCCAGATATCAACGACTATTTCAAAAAGTTTTCGA 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22 GlyaentpGlnTyLyserleuaspValaenValaenileGluGlnaenPhelle--- 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 180 CCTGAA-----CCAGATACAGCCTTTCTTGGTGAAGAAATGTTGGAAGAA 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 ProaspIleaspSerAlaValaArgilelleProValaenTyAspSeraspProlysleu 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 228 GGAGCTTATTTATTTAGTGAATCTTCTGCTGCAAAATGTATCCAGGTAACGCTC 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 AenSerGlnleuTyThValGluMetThrIleProAlaGlyValSerAlaValyslle 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 288 TACCT-----GTTATGATGAAGATTATGGGTAGAGCACTAGTAATACCGCT 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81 ValProthrAaspSerleuThserSerGlyGlnGlnileGlyLeuValaenValaen 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 339 GATGCTTCCCATCAATATCTACCAATGTGTGATGAGAA-----GGGAAAAAATG 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 AenProaspGlnaenMetAenTyThxIleArglyAspSerGlyAlaGlylyserhete 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 393 TTAAGAGATCATGTGCA---GAGTTACACCTATCAACAATATCTTTAAACGCTG 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AlaGlyGlnlyGlySerPheSerValylsGluAenThserTyThPheSerAlalle 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 450 AATATATACGCGGGAA-----AAAAAATATCTCTGGAATATATACATG 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 141 ---TyThrGlyGlyTyThProaenSerGlyTySerSerGlyThTyAlaGlyHis 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 501 GTTATGTTGTTACTATGTAAC 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 LeuThrValSerPheTySerAsn 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 4

FB82_ECOLI STANDARD; PRT; 167 AA.

AC P53511;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE CS6 fimbrial subunit B precursor.

GN Name=cs6B;

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=E8775;

RA Gasestra W., Warren R., Boedeker E.C.;

RA Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not removed.

CC EMBL; U04846; AAB51362.1; -; Unassigned_DNA.

KW Fimbria; Signal.

FT SIGNAL 1 21

SQ SEQUENCE 167 AA; 17930 MW; B8DD041D3BFBDCB CRC64;

Alignment Scores:

Pred. No.: 0.0187 Length: 167
Score: 118.00 Matches: 49
Percent Similarity: 42.1% Conservative: 26
Best Local Similarity: 27.5% Mismatches: 67
Query Match: 12.6% Indels: 36
DB: 1 Gaps: 9

US-10-768-093-4 (1-527) x FB82_ECOLI (1-167)

```

OY 63 ATGAGAAACAATGTTTATTTCTAATCTTCTGTCATTCGAGCCATCCAGCA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 LeuylsylvilletteSerAlalleleuilealea-----GlyThrSerGlyValVal 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 123 GAATACGACT-----AAAACCTCCCATCAACGACTATTTCAAAAAGTTT 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20 AenAlaGlyaentpGlnTyLyserleuaspValaenValaenileGluGlnaenPh 39
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 174 TTTCACCTGAA-----CCAGATACAGCCTTTCTTGGTGAAGAAATGTTGGA 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40 Ile---ProaspIleaspSerAlaValaArgilelleProValaenTyAspSeraspPro 58
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 222 AAGAGAGAGCTTTATTTATTTAGTGAATCTTCTGCTGCAAAATGTATCCAGTA 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 59 LysleuaspSerGlnleuTyThValGluMetThrIleProAlaGlyValSerAlaVal 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 282 ACGGCTACCT-----GTTATGATGAAGATTATGGGTAGAGCACTAGTAAT 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 79 LysIleAlaProthrAaspSerleuThserSerGlyGlnGlnileGlyLyLeuValaen 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 333 ACCGCTGATGCTCCCATCAATATCTACCAATGTGTGATGAGAAAGGAAAAAATG 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99 ValaenAenProaspGlnaenMetAenTyThxIle-----GAGTTACA 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 393 TTAAGATCATGTGCA-----GAGTTACA 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 ArglyAspSerGlyAlaGlyAsnPhenAlaGlyGlnlySerPheProValys 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 420 CCTAATCAACAATAATCTTTAAAGCGCTGAATATCTACGCGGGAA-----AAA 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 131 GluAenThserTyThPheSerAlalle---TyThrGlyGlyTyThProaenSer 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 471 AAAATATCTCTGGAATATATACATCAAGCTTATGTTGTTACTATGTAAC 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 GlyTySerSerGlyThTyAlaGlyAsnleuThrValSerPheTySerAsn 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

Q54US8_DICDI PRELIMINARY; PRT; 440 AA.

AC Q54US8;

DT 13-SEP-2005 (TRENDEL 31, Created)

DT 13-SEP-2005 (TRENDEL 31, Last sequence update)

DT 13-SEP-2005 (TRENDEL 31, Last annotation update)

DE Hypothetical protein.

GN ORFNames=DD80215209;

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Eukaryota; Dictyostelida; Dictyostelium.

OX NCBI_TaxID=44689;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=AX4;

RA Eichinger L., Pachepat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Bertman M., Song J., Olsen R., Szafinski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,

OK NCBI_TaxID=1313;
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=53/62, 688/6, and 7904/39;
 RC Bentley S.D., Aarensen D., Mavroidi A., Saunders D., Rabinowitch E.,
 RA Collins M., Donaghe K., Harris D., Kallio M.S., Murphy L.,
 RA Quail M.A., Samuel G., Skovsted I.C., Barrell B.G., Reeves P.,
 RA Parkhill J., Spratt B.G.;
 RT "Genetic analysis of the capsular biosynthetic locus from all 90
 serotypes of *Streptococcus pneumoniae*."
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, CR931665; CAI33391.1; -; Genomic_DNA.
 DR EMBL, CR931666; CAI33414.1; -; Genomic_DNA.
 DR EMBL, CR931664; CAI33368.1; -; Genomic_DNA.
 KW Transferase.
 SQ SEQUENCE 318 AA; 37207 MW; 26ADD45FBFBED25 CRC64;
 Alignment Scores:
 Pred. No.: 8.54 Length: 318
 Score: 90.50 Matches: 43
 Percent Similarity: 42.84 Conservative: 19
 Best Local Similarity: 29.74 Mismatches: 58
 Query Match: 9.64 Indels: 25
 DB: 2 Gaps: 6
 US-10-768-093-4 (1-527) x Q4K1L8_STRPN (1-318)
 QY 60 TATATGAGAAACATGTTTAAATCTA-----ATT 92
 DB 17 TyrLeuLysLysSerLleGluSerValleuAnGlnThyTyrGlnAenLleGluLleu 36
 QY 93 CTTCCTTATTCGCGAGCCATGCCAGAACAGAAATAGCGACTTAAAACTCCAGTATCA 152
 DB 37 LeuValaAspGlySer-----ThraSerSerGlyLys----- 48
 QY 153 ACGACTATTTCAAAAGTTTTCGACCTGACCAACGAAATACAGCTTCTTTGGTGA 212
 DB 49 -----IleCyGluSerPheSerLysValaAspProArgLleArgValaPheHisLysGlu 66
 QY 213 AATGTGGAAGAGAGAGAGCTTATTTAGTGTGGAAGTAACTTCTGAAATGTA 272
 DB 67 AenGlyGlyLeuSerAspAlaArgAenPheGlyLleGluGlnMelLysGlyGlnTyrVal 86
 QY 273 TCCAGGTACGCTGTAACCTGTTATGATGAGATTA-----GGGTAGAGAGACTA 326
 DB 87 AlaPheLleAspSerAspAspTyrLleSerLysAspTyrValTrrLysLeuTyrSerSer 106
 QY 327 GTAATACCGCTGATGCTTCCCAATCAATATC--TACCAGATTGTGATGAGAAAGG 383
 DB 107 IleLysAenAenAspSerGluValSerLleCySerPheLeuLeuValaAspGluLysGly 126
 QY 384 AAAAAAATG-----TTAAAGATCATGTGCGAGAGTTACACCTAATCAACAATA 434
 DB 127 GluLysLleLysAspGluLeuLeuAspSerGlyLysLleCySerLeuThrgLysGlnIle 146
 QY 435 ACTTTTAAAGCGCTG 449
 DB 147 LeuGluLysValleu 151
 RESULT 8
 Q4K1O5_STRPN PRELIMINARY; PRT; 318 AA.
 AC Q4K1O5;
 DT 13-SEP-2005 (TREMBLrel. 31, Created)
 DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
 DE Putative glycosyl transferase.
 GN ORFNames=SPC15A_0013;
 OS *Streptococcus pneumoniae*.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus;
 OK NCBI_TaxID=1313;
 RN [1]

RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=389/39;
 RA Bentley S.D., Aarensen D., Mavroidi A., Saunders D., Rabinowitch E.,
 RA Collins M., Donaghe K., Harris D., Kallio M.S., Murphy L.,
 RA Quail M.A., Samuel G., Skovsted I.C., Barrell B.G., Reeves P.,
 RA Parkhill J., Spratt B.G.;
 RT "Genetic analysis of the capsular biosynthetic locus from all 90
 serotypes of *Streptococcus pneumoniae*."
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL, CR931663; CAI33345.1; -; Genomic_DNA.
 KW Transferase.
 SQ SEQUENCE 318 AA; 37209 MW; 0190E7D08D6AAE33 CRC64;
 Alignment Scores:
 Pred. No.: 10.7 Length: 318
 Score: 89.50 Matches: 42
 Percent Similarity: 42.84 Conservative: 20
 Best Local Similarity: 29.04 Mismatches: 58
 Query Match: 9.54 Indels: 25
 DB: 2 Gaps: 6
 US-10-768-093-4 (1-527) x Q4K1O5_STRPN (1-318)
 QY 60 TATATGAGAAACATGTTTAAATCTA-----ATT 92
 DB 17 TyrLeuLysLysSerLleGluSerValleuAnGlnThyTyrGlnAenLleGluLleu 36
 QY 93 CTTCCTTATTCGCGAGCCATGCCAGAACAGAAATAGCGACTTAAAACTCCAGTATCA 152
 DB 37 LeuValaAspGlySer-----ThraSerSerGlyLys----- 48
 QY 153 ACGACTATTTCAAAAGTTTTCGACCTGACCAACGAAATACAGCTTCTTTGGTGA 212
 DB 49 -----IleCyGluSerPheSerLysValaAspProArgLleArgValaPheHisLysGlu 66
 QY 213 AATGTGGAAGAGAGAGCTTATTTAGTGTGGAAGTAACTTCTGAAATGTA 272
 DB 67 AenGlyGlyLeuSerAspAlaArgAenPheGlyLleGluGlnMelLysGlyGlnTyrVal 86
 QY 273 TCCAGGTACGCTGTAACCTGTTATGATGAGATTA-----GGGTAGAGAGACTA 326
 DB 87 AlaPheLleAspSerAspAspTyrLleSerLysAspTyrValTrrLysLeuTyrSerSer 106
 QY 327 GTAATACCGCTGATGCTTCCCAATCAATATC--TACCAGATTGTGATGAGAAAGG 383
 DB 107 IleLysAenAenAspSerGluValSerLleCySerPheLeuLeuValaAspGluLysGly 126
 QY 384 AAAAAAATG-----TTAAAGATCATGTGCGAGAGTTACACCTAATCAACAATA 434
 DB 127 GluLysLleLysAspGluLeuLeuAspSerGlyLysLleCySerLeuThrgLysGlnIle 146
 QY 435 ACTTTTAAAGCGCTG 449
 DB 147 LeuGluLysValleu 151
 RESULT 9
 Q37318_HABPV PRELIMINARY; PRT; 1007 AA.
 AC Q37318;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Spheroidin.
 OS *Heliothis armigera entomopoxvirus* (HABPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae.
 OK NCBI_TaxID=10290;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=98062131; PubMed=9400959;
 RA Srikantha A., Osborne R.J., Dall D.J.;
 RT "Mapping of the *Heliothis armigera entomopoxvirus* (HABPV) genome, and
 analysis of genes encoding the HABPV spheroidin and nucleoside
 triphosphate phosphohydrolase I proteins.";

RL J. Gen. Virol. 78:3115-3123(1997).
 RN (2)
 RP NUCLEOTIDE SEQUENCE.
 RA Srikenantha A., Osborne R.J., Dail D.J.
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF019224; AAB96623.1; -; Genomic_DNA.
 DR InterPro; IPR008843; Spheroidin.
 DR Pfam; PF05541; Spheroidin; 1.
 SO SEQUENCE 1007 AA; 11538 MW; 3352847F56586D8E CRC64;

 Alignment Scores:
 Pred. No.: 14.5 Length: 1007
 Score: 88.50 Matches: 33
 Percent Similarity: 42.2% Conservative: 29
 Best Local Similarity: 22.4% Mismatches: 46
 Query Match: 9.4% Indels: 39
 DB: 2 Gaps: 6

 US-10-768-093-4 (1-527) x O37318_HABPV (1-1007)
 QY 54 TGGTGTATATGAGAAACAAATTGGTTAATCTTATCTTCCTCATTCGGCAGCCAT 113
 DB 99 TTPlySTyValSerArg-----LeuLeuLeuAlaSerVal---SerHis 112
 QY 114 GCCAGAACAGAAATAGCAGCTAAACCTTCCAGATCAACGACTATTTCAAAAAGTTT 173
 DB 113 AapAenVallyTyryLyseuAlaenTyryLyseu-----ThreLeuApenlySHis 130
 QY 174 TTTCGACCTGAAACGAAATACAGCCT-----TCITTTGGTAAATGTGGAAG 224
 DB 131 lIeLySProLyshIleIleAspInProLeuPheIleTyryPheValAspAspleuInHis 150
 QY 225 GAAGAGCTTTA-----TTATTAGTGGACTTAATCTGCTGGAATGTA 272
 DB 151 TyryGlyLeuIleThryIsegluAenIleTyryAsnAspAsnleuInValAsnGluAspAla 170
 QY 273 TCCCAAGTACGCTTACCTCTGTTAT----- 299
 DB 171 SerPheIleThryValPheProGlnTyryAlaTyryIleHisIleGlyArgLyseValTyryIle 190
 QY 300 -----GATGAAATATATGGTTAGGA 320
 DB 191 AenGluLyseSerThryPheAspValThryThraSpAlaThryAsnIleAsnIleAsnPheAsn 210
 QY 321 CGACTAGTAATATACCGCTAGCTGCTCCCAATCAATATATCAGCAGATGTTGATGAGAA 380
 DB 211 LyseSerValAsnIleAlaValSerPheLeuApeIleTyryGlnValAsnAsnAsnGln 230
 QY 381 GGGAAAAAATGTTAAAGAT 401
 DB 231 GluLyseGluLeuLeuLyseAsp 237

 RESULT 10
 Q54ED6_DICDI PRELIMINARY; PRT; 2542 AA.
 AC 054ED6;
 DT 13-SEP-2005 (TReMBLrel. 31, Created)
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=DOB0219680;
 GN Dicyostellium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dicyostelidia; Dicyostelium.
 OX NCBI_TaxId=44689;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=AX4;
 RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
 RA Sugeng R., Berthman M., Song J., Olsen R., Szafanski K., Xu Q.,
 RA Tungsang B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
 RA Benlier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
 RA Plicher K., Chen G., Saunders D., Sodergren E., Davis P.,
 RA Kernionou A., Nie X., Hall N., Anjard C., Hemphill L., Baeson N.,

RA Fairbrother P., Desany B., Just E., Morio T., Rose R., Churcher C.,
 RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
 RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
 RA Hauser H., James K., Oulles M., Mohan M.B., Saito T., Buchrieser C.,
 RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
 RA Louised H., Mungall K., Oliver K., Price C., Qual M.A., Sanders M.,
 RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D.,
 RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
 RA Sugano S., White B., Walker D., Woodward J., Winkler T., Tanaka Y.,
 RA Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
 RA Chisolem R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
 RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,
 RT The genome of the social amoeba Dicyostellium discoideum."
 RL Nature 0:0-0(2005).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAF10100257; BAF61826.1; -; Genomic_DNA.
 KW Hypothetical protein.
 SO SEQUENCE 2542 AA; 28853 MW; D28121989617A16 CRC64;

 Alignment Scores:
 Pred. No.: 17.4 Length: 2542
 Score: 88.00 Matches: 38
 Percent Similarity: 39.2% Conservative: 27
 Best Local Similarity: 22.9% Mismatches: 59
 Query Match: 9.4% Indels: 42
 DB: 2 Gaps: 6

 US-10-768-093-4 (1-527) x Q54ED6_DICDI (1-2542)
 QY 72 ACAATTGGTTAATCTTAATCTTCTTCATTCGGCAGCCATGCCAGAACAGAAATAGCG 131
 DB 2357 SerIleGlyLeuProThryIleCySerThraenTyryGlyLeuIleGluSerThryGlyPheVal 2376
 QY 132 ACTAAACCTTCCAGTATCAACGACTATTTCAAAAAGTTTTCACCT----- 182
 DB 2377 SerArgAsnGluSerValAlaIleAlaLeuLeuSerGlyGluGlyPheIleProIleSerAla 2396
 QY 183 -----GAAACGAAATACAGCCTCTTCTTGTGGAATGTGGA 221
 DB 2397 AsnLeuIleLeuGlyThryLeuAspleuInleuGln-----AsnGlnAlaGln 2412
 QY 222 AAGGAAGAGCTTATTAATTAAGTGAATTAATCTTCTGAAAATGATCCAGGTA 281
 DB 2413 SerSerAsnLeuIleLeuSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 2432
 QY 282 ACGGCTACCCGTTTATGATGAAATATATGCGTTAGACAGCATGTAATACCGCTGAT 341.
 DB 2433 SerLeuIleSerLysePheAsp----- 2439
 QY 342 GCTCCCAATCAATATCTACCAAGT-----GTTGATGAGAAAGGAAAAA 389
 DB 2440 -----TyryGlnIleAsnIleAsnGluIleAsnGluIleAsnGluIleAsnGluIleAsn 2453
 QY 390 ATGTTAAAGATCATGTGACAGGTTACACCTTAATCAACAATATCTTTAAAGCGCTG 449
 DB 2454 LeuLeuLyseAspAspAsnValGluIleThryValAspGlnIleIleThryPheLyseSer 2473
 QY 450 AATTATCTACGCGGGGAA-----AAAAAATATCTCTGGAATATATAGATAGATTATG 506
 DB 2474 GluIleLeuSerThryAspIleLeuLyseAsn-----LyseAspIleIleLeu 2489
 QY 507 GTTGTTACTATGTAAC 524
 DB 2490 ValAspTyryGlyIleAsp 2495

 RESULT 11
 Q72P30_LEPIC
 ID Q72P30 LEPIE PRELIMINARY; PRT; 918 AA.
 AC Q72P30;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Membrane carboxypeptidase.
GN OrderedLocustNames=LIIC12646;
OS Leptocystis interrogans (serogroup Icterohaemorrhagiae / serovar
OS Copenhagen).
OC Bacteriia; Spirochaetes; Spirochaetiales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=44275;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=F10crz LI-130;
RX PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L.T., Monteiro-Vitorello C.B.,
RA Ho P.L., Haake D.A., Verlyovski-Almeida S., Hartkeerl R.A.,
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrier H.,
RA Coutinho L.L., Degrange W.M., Dellacostin O.A., El-Dorry H.,
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gambellini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harkava R.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T.,
RA Kuramae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schrieffer A.,
RA Siqueira W.J., Sommer P., Teal S.M., Simpson A.J.G., Ferro J.A.,
RA Canagaro L.E.A., Kitajima U.P., Setubal J.C., Van Sluys M.A.;
RT "Comparative genomics of two Leptocystis interrogans serovars reveals
RT novel insights into physiology and pathogenesis.";
RL J. Bacteriol. 186:2164-2172(2004).
DR EMBL: A601297; A657106.1; -; Genomic DNA.
DR DR GO: 0009274; C:cell wall (barnu Bacteriia); IEA.
DR GO: 00004180; F:carboxypeptidase activity; IEA.
DR GO: 0003824; F:catalytic activity; IEA.
DR GO: 0008658; F:penicillin binding; IEA.
DR GO: 0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro: IPR001264; Glyco_trans_51.
DR InterPro: IPR001460; Genc1_bind_cpptc.
DR Pfam: PF00912; Transglyl; 1.
DR Pfam: PF00905; Transpeptidase; 1.
DR Prodom: PD001895; Glyco_trans_51; 1.
KW Carboxypeptidase; Complete proteome.
QC Sequence 918 AA; 102807 MW; 2D5E521BC396960 CRC64;

Alignment Scores:			
Pred. No.:	20.1	Length:	918
Score:	87.00	Matches:	37
Percent Similarity:	43.4%	Conservative:	25
Best Local Similarity:	25.9%	Mismatches:	57
Query Match:	9.3%	Indels:	24
DB:	2	Gaps:	6
US-10-768-093-4 (1-527) × Q72P30.LBPC (1-918)			
QY	132 ACTAAAACTTCCAGTATCAACGACTATTTCAAAAAGTTTTCGACCTGACACGACA	191	
Db	478 ThrGlyGlyPheSerIeIyThrThrValSerGluProValGlnIaGluLeuProlYs	497	
QY	192 ATACGACCTTCCTTTGGTGCMAAATGTTGGAAAGGAGGACCTTATTATTAGTGCAGAC	251	
Db	498 ILeValIyAsnTrpValAspAsnValGlnIyAsnGly--LeuValIArgIySthArg	516	
QY	252 TTAACCTGTCCGAAATGTATCCCGACGATACGGGTCTACCCCTGTTATGATGAAATATAT	311	
Db	517 LeuThrAspAsnIyLeuAsnSerSerGluThrAlaValIAspArgArgIyIleGlnAspLeu	536	
QY	312 -----GGGTTTAGACGACCTAGTAATACC-----GCTAGTGCCTCC	347	
Db	537 SerProAlaLeuGluLeuPheIleAspThrAspSerPheGlyGlnIAsnAsnGluSerGly	556	
QY	348 CAATCAATAATCTACACAGATTGTCATGTGAGAAAGGAAAAATGTTAAAGATCATGGT	407	
Db	557 LeuGlnIValAlaLeuValAlaValAspProSerThrGlyGluIleLeuIleuMetHisGly	576	
QY	408 GCAGAGGTTACACCTTAATCAACAAATACTTTAAAGC-----CTGAATTATACT	458	
Db	577 GlySer-----GluPheIyAlaAspAsnGlnIleuAspArgIthr	589	

```

QY      459 AGCGGGGAAAAAAA-----ATACCTCGGAATATTAAGCATCAGTT 503
           |||:::|
Db      590 ThAlaMeLrYgLnThrClYseSerIleYsProlLeuYrSerAlaIaIle 603
           |||:::|
QY      504 ATGGTTGGT 512
           |||
Db      610 GluThrClY 612

RESULT 12
Q8F7D7 LEPIN
ID      Q8F7D7 LEPIN PRELIMINARY; PRT; 918 AA.
AC      Q8F7D7;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Penicillin-binding protein 1A (EC 3.4.-.-) (EC 2.4.2.-).
GN      OederelococcusNames=IA1009;
OS      Leptospira interrogans.
OC      Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX      NCBI_taxid=173;

```

NP COLLECTIVE SEQUENCE.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
 RX MEDLINE=22558143; PubMed=12712204; DOI=10.1038/nature01597;
 RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
 RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-O., Jia J., Tu Y.-F.,
 RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
 RA Zhang Y., Zhu G.-F., Wan M., Huang X.-L., Qian Z., Wang S.-Y., Ma W.,
 RA Yao Z.-Y., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
 RA Saint Girones I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
 RA Xu J.-G., Zhao G.-P.;
 RT "unique physiological and pathogenic features of *Leptospira*
 RT interrogans revealed by whole-genome sequencing.",
 RL Nature 422:888-893(2003).
 EMBL; AEO11285; AAN48208.1; -; Genomic DNA.
 DR GO: GO:0009274; C:cell wall (sensu Bacteria); IEA.
 DR GO: GO:0003824; P:catalytic activity; IEA.
 DR GO: GO:0006658; P:penicillin binding; IEA.
 DR GO: GO:0009252; P:peptidoglycan biosynthesis; IEA.
 DR InterPro: IPR001264; Glyco_trans_51.
 DR InterPro: IPR001460; Pencil_bind_type.
 DR Pfam: PF00912; Transgly_1.
 DR Pfam: PF00905; Transpeptidase; 1.
 DR Prodom: PD001895; Glyco_trans_51; 1.
 KW complete proteome.
 QO SEQUENCE 918 AA; 102818 MW; 5E8A6FD48DC58326 CRC64;

Alignment Scores:		20.1	Length:	918
Pred. No.:		87.00	Matches:	37
Score:		43.4%	Conservative:	25
Percent Similarity:		25.9%	Mismatches:	57
Best Local Similarity:		9.3%	Indels:	24
Query Match:			Gaps:	6
DB:				

US-10-768-093-4 (1-527) x Q8P7D7_LBPIN (1-918)	
QY 132 ACTAAAAACCTTCGACATCAACGACGACTATTTTCAAAAAGTTTTCCTGACCTGAACACGA 191	
DB 478 ThGlyGlyPhseSerIeTyrThrThrValSerGluProValGlnAlaGluLeuProIys 497	
QY 192 ATACAGCCCTCTTCTTGCTGGAAGAAATGTTGGAAAGGAGAGCGTTATTATTAGTGTGAAC 251	
DB 498 IleValIysAsnTyrValAspAsnValGlnIysAsnGly---LeuValArgIysThrArg 516	
QY 252 TTAAGTCTTCGAAAATGTATCCAGGTAAACGGTCTACCTGTTTATGATGAAGATTAT 311	
DB 517 LeuThrAspAsnIlyAsnSerSerGluThrAlaValAlaPheArgArgTyrIleGlnAspLeu 536	
QY 312 -----GGGTTAGAGACGACTAGTAATAAC-----GCTGAGCTTCC 347	
DB 537 SerProAlaLeuGluLeuPheIleAspThrAspSerPheGlyValGlnAsnGluSerGly 556	

US-10-768-093-4 (1-527) x Q8IAS0_PLAF7 (1-573)

QY 28 AATAAAATGTTATGCAATATTAATGTTATGCAAGAAAATGTTATTC 87
 DB 187 AenilunhensSerAspAsnGluAsnAsn-LeuAspAsnGluAsnAsnLeuAspAsnGlu 206
 QY 88 TAATCTCTGCTTCATTCGGCAGCCATGCCAGAACGAAATAGCG-----ACTAAAACT 141
 DB 206 uAsnAsn-SerAspAsnSerSerIleGluAsnPheIleAlaLeuGluAsnLysAsnA 226
 QY 142 TCCCAATATCAACACTATTTCAAAAATTTTTCACCTGAAACACAGATACAGCCTT 201
 DB 226 lArhValGluGlnThrLysGluAsnIlePheLeuValProLeuLysHileuAsnAsp 246
 QY 202 CTTTGGTGAATGTTGAAG 261
 DB 246 er-----GlnPheValGlyLeu-----LeuLeuValGly-----ThrProP 258
 QY 262 CTGAAAAATGATCCAGGTAACGGTCTACCTGTTATGATGAAGATTAAGGTTAGAC 321
 DB 258 roGln-----ThrValTyrProIlePheAsp----- 266
 QY 322 GACTAGTAATACCGCTGATGCTCCCAATCATATATACCAATGTTGATGAGAAG 381
 DB 267 -----ThrGlySerThrAsnValTyrValValThrAlaCysGluGluGln 283
 QY 382 GAAAAAATGTTAAAGATCATGTCAGAGAGTAACTAATCAACAATTAATCTTTT 440
 DB 283 erCysLysLysValAlaGlyArgTyr-----AspProAsnLysSerLysThrPheA 299
 QY 441 -----AAACCGCTGAATTAATTAAGCGGGGAAAAAAATATCTCTG 483
 DB 299 rGATGSerPheIleGluLysAsnLeuHisIleValPheGlySerGlySerIleSerGlyS 319
 QY 484 GAATATATACGATCAGCTATGCTGTTACTATGTA 521
 DB 319 erValGlyThrAspThrPheMetLeuGlyLysHisLeu 331

RESULT 15

Q9M2C6_ARATH PRELIMINARY; PRT; 718 AA.

AC Q9M2C6;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein T20K12.280.
 GN Name=T20K12.280;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA De Haen M., Maere A.C., Grivell L.A., Newes H.W., Lemcke K.,
 RA Mayer K.F.X., Quelier F., Salanoubat M.,
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL137898; CAB71068.1; -; Genomic_DNA.
 DR PIR; T47930; T47930.
 KW Hypothetical protein.
 SO SEQUENCE 718 AA; 80730 MW; F3DCA7B9379AEB25 CRC64;

Alignment Scores:

Pred. No.: 27.4 Length: 718
 Score: 85.50 Matches: 37
 Percent Similarity: 41.9% Conservative: 28
 Best Local Similarity: 23.9% Mismatches: 73
 Query Match: 9.1% Indels: 17
 DB: 2 Gaps: 2

US-10-768-093-4 (1-527) x Q9M2C6_ARATH (1-718)

QY 69 AAAAAATGTTTAAATTTCTAATCTTCTTCATTCGGCAGCCATGCCAGAACAGAAATA 128
 DB 334 ArgIleuGlyLysIleLeuSerLeuProGluPheCysSerProAlaAspSerProArg 353
 QY 129 GCGACTAAAACTCCAGATATCAACACTATTTCAAAAATTTTTCACCTGAACCA 188
 DB 354 LeuIleProAlaHisAspLeuValSerThrLeuSerGlnThrGluGlnProGluIle 373
 QY 189 CGAATACAGCCTTCT-----TTTGTGAATATGTTGAAG 224
 DB 374 LeuGlnThrProGluThrSerSerAlaThrAsnAspLeuIleAspGluAspSerAspLys 393
 QY 225 GAAGAGCTTATATTAATTAATGTAACCTTAATCTGTAATAATGTAATCCAGTAAAG 284
 DB 394 AspAspAspThrLeuPheThrIleAspValSerValProArgAspTyrGlyAsnGluThr 413
 QY 285 GTCTACCTGTTTATGATGAAGATTATGAGTAAAGACTAGTAATACCGCTGATGCT 344
 DB 414 GluAsnIleAspAsnAspGluGlnSerGluIleAspProLeuSerGluThrCysSerSer 433
 QY 345 TCCCAATCAATTAATCTACAGATTGTTGATGAGAAAGGAAAAAATGTTAAAGATCAT 404
 DB 434 SerValSerArgGluValGluAsnValAspIleAspValGluLysGluMetLeuAsnHis 453
 QY 405 GGTGCAGAG-----GTTACACCTTAATCAACAATAACT 437
 DB 454 GlnAlaHisSerProAlaSerProLeuGlnSerProThrThrCysSerValArgLeuThr 473
 QY 438 TTTAAAGCGCTGAATTAATTAATGAGCGGAAAAAAATATATCTCT 482
 DB 474 GluCysLysGluThrAlaThrAspAlaGlnGlyLysLeuSerPro 488

Search completed: April 10, 2006, 10:06:14
 Job time : 186.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW nucleic - protein search, using frame_plus_n2p model

Run on: April 10, 2006, 10:06:30 / Search time 6.7 Seconds
(without alignments)
1300.599 Million cell updates/sec

Title: US-10-768-093-4
Perfect score: 939
Sequence: 1 TTGCACATTCAGCATGTTA.....TTGTTACTATGTAACATA 527

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 572060 seqs, 82655679 residues
Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model-DEV=xlp
-Q=/abse/ABSSWB_spool/US10768093/runat_10042006_090307_4612/app_query.fasta_1
-DB=Issued_Patents_AA-QEWT=faeacan-SUFFIX=n2p.ra1-MINMATCH=0.1-LOOPCL=0
-LOOPEXT=0-UNITS=bits-START=1-END=1-MATRIX=blosum62-TRANS=human40.cdi
-LIST=45-DOCALLIGN=200-THR SCORE=pct-THR MAX=100-THR MIN=0-ALIGN=15
-MODE=local-OUTPM=pcio-NOR=ext-HEAPSIZ=500-MINLEN=0-MAXLEN=200000000
-HOST=abse02p-USER=US10768093@cgn.1.1.71@runat_10042006_090307_4612-NCPU=6
-ICPU=3-NO_MAP-NEG SCORES=0-WAIT-DSBLOCK=100-LOGLOG-DEV TIMEOUT=120
-WARN TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6-FGAPEXT=7
-YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCUTS_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	9.2	378	2	US-09-107-532A-4742 Sequence 4742, Ap
2	84	8.9	740	2	US-09-323-872A-25 Sequence 30, Appl
3	84	8.9	740	2	US-09-072-433-30 Sequence 2, Appl
4	82	8.7	330	2	US-09-248-796A-14853 Sequence 14853, A
5	81	8.6	768	2	US-09-134-000C-6467 Sequence 6467, Ap
6	77	8.2	345	2	US-09-538-092-314 Sequence 314, App
7	77	8.2	483	2	US-09-252-991A-21015 Sequence 21015, A
8	76.5	8.1	186	2	US-09-543-681A-6955 Sequence 6955, Ap
9	76.5	8.1	308	2	US-09-248-796A-18392 Sequence 18392, A
10	76.5	8.1	1076	2	US-09-134-000C-4037 Sequence 4037, Ap
11	76	8.1	218	2	US-09-328-352-6886 Sequence 6886, Ap
12	76	8.1	527	2	US-09-252-991A-23709 Sequence 23709, A

13	75.5	8.0	603	2	US-09-902-540-11397 Sequence 11397, A
14	75.5	8.0	1184	2	US-09-541-782-2 Sequence 2, Appl
15	75.5	8.0	1184	2	US-09-723-820-2 Sequence 2, Appl
16	75.5	8.0	1184	2	US-10-270-085-2 Sequence 2, Appl
17	75	8.0	130	2	US-09-134-000C-5020 Sequence 5020, Ap
18	74.5	7.9	400	2	US-09-710-279-2774 Sequence 2774, Ap
19	74.5	7.9	497	2	US-09-134-001C-5114 Sequence 5114, Ap
20	74	7.9	1070	1	US-08-913-477-21 Sequence 21, Appl
21	74	7.9	302	2	US-09-328-352-7673 Sequence 7673, Ap
22	74	7.9	467	2	US-09-248-796A-17536 Sequence 23, Appl
23	74	7.9	501	1	US-08-913-477-23 Sequence 23, Appl
24	73.5	7.8	125	2	US-09-328-352-5033 Sequence 5033, Ap
25	73.5	7.8	584	2	US-09-107-532A-4564 Sequence 4564, Ap
26	73.5	7.8	1006	2	US-09-710-279-154 Sequence 154, App
27	73.5	7.8	1806	2	US-09-919-497-56 Sequence 56, Appl
28	72.5	7.7	465	2	US-09-769-787-164 Sequence 164, App
29	72	7.7	1005	2	US-09-134-000C-4976 Sequence 4976, Ap
30	72	7.7	1005	2	US-09-949-016-6079 Sequence 6079, Ap
31	72	7.7	1008	2	US-09-949-016-7359 Sequence 7359, Ap
32	71.5	7.6	84	2	US-09-270-767-37692 Sequence 37692, A
33	71.5	7.6	84	2	US-09-270-767-52909 Sequence 52909, A
34	71.5	7.6	375	2	US-09-270-767-11678 Sequence 31678, A
35	71.5	7.6	375	2	US-09-270-767-46895 Sequence 46895, A
36	71.5	7.6	399	2	US-09-583-110-2754 Sequence 2754, Ap
37	71.5	7.6	560	2	US-08-983-045-4 Sequence 4, Appl
38	71.5	7.6	1467	2	US-09-134-000C-6740 Sequence 6740, Ap
39	71	7.6	289	2	US-09-107-532A-7169 Sequence 7169, Ap
40	70.5	7.5	305	2	US-09-328-352-6069 Sequence 6069, Ap
41	70.5	7.5	323	2	US-09-538-092-213 Sequence 213, App
42	70.5	7.5	357	2	US-09-103-331-41 Sequence 41, Appl
43	70.5	7.5	357	2	US-09-631-594-50 Sequence 50, Appl
44	70.5	7.5	452	2	US-09-538-092-371 Sequence 371, App
45	70.5	7.5	542	2	US-09-107-532A-4858 Sequence 4858, Ap

ALIGNMENTS

RESULT 1
US-09-107-532A-4742
Sequence 4742, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998.
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4742:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...378
SEQUENCE DESCRIPTION: SEQ ID NO: 4742:
US-09-107-532A-4742

Alignment Scores:
Pred. No.: 0.0337 Length: 378
Score: 86.00 Matches: 39
Percent Similarity: 42.7% Conservative: 25
Best Local Similarity: 26.0% Mismatches: 56
Query Match: 9.2% Indels: 30
Gaps: 8

US-10-768-093-4 (1-527) x US-09-107-532A-4742 (1-378)

QY 90 ATTCTGCTTCATTCGCGAGCATGCGAGACAGAAATAGCGACTAAACTCCAGTA 149
:::|||||:::|||||:::|||||:::|||||
DB 158 VALLEALALALAGLYALAGLYTHRGINGLYASPTY-THRGlyThrsrhe----- 175
150 TCAAGACTATTTCAGAAAGTTTTCGACCTGAAACGAGATACAGCTTCTTTGCT 209
176 ---ThrsrlelglYHlglYArGPhetrProGluAlglYAlglInProAlallegly 194
QY 210 GAAATGTTGGAAG---GAGGAGCTTATATTATTAGTGAACCTGACTGCTGAA 266
:::|||||:::|||||:::|||||:::|||||
DB 195 Lye---ValGlyLylThrgInglInValGIngluAlalYValgluVallleuProglu 213
267 AATGATCCAGTAAAGGCTTAC-----GAGATTAAGCTTACGAGACTAGTAAT 296
:::|||||:::|||||:::|||||:::|||||
DB 214 ThrleGlulYglInVallelglInAlameArGserAlahlsProtyrGlulInProAla 233
297 TATGAT-----GAGATTAAGCTTACGAGACTAGTAAT 332
|||||:::|||||:::|||||:::|||||
QY 234 TyArPleuPheAlalaleaPgluProValGlumecPheGlyLeuclYArGVal----- 251
333 ACCGCTGATGCTCCATCAATTAATCTACAGATTGTGAGAGAAAGGAAATATG 392
252 ---GlygluLeuProGInglulIlePhelelgluAlaPheValgluInVallysgluAla 270
393 TTAAAGATCATGCTGCGAGAGTTTACA-----CCTAATCAACAAATTACTTTTAA-- 443
:::|||||:::|||||:::|||||:::|||||
DB 271 PheGInleuAspGlyleuArGleValglInProlyAsnaAlalYserseVallylsArG 290
444 ---GGCTGAATTATTAATCGCGGAAAAA 470
DB 291 lleaAlalecYeglyclYserGlygluLys 300

RESULT 2
US-09-323-872A-25
Sequence 25, Application US/09323872A
Patent No. 6395539
GENERAL INFORMATION:
APPLICANT: Coschigano, Peter
TITLE OF INVENTION: Compositions and Methods for Bioremediation
FILE REFERENCE: OHU-03640
CURRENT APPLICATION NUMBER: US/09/323, 872A
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/072, 433
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.0
SEQ ID NO 25
LENGTH: 740

TYPE: PRT
ORGANISM: Clostridium pasteurianum
US-09-323-872A-25

Alignment Scores:
Pred. No.: 0.0849 Length: 740
Score: 84.00 Matches: 39
Percent Similarity: 38.6% Conservative: 25
Best Local Similarity: 23.5% Mismatches: 68
Query Match: 8.9% Indels: 34
Gaps: 7

US-10-768-093-4 (1-527) x US-09-323-872A-25 (1-740)

QY 93 CTGCTTCATTCGCGAGCATGCGAGACAGAAATAGCGACTAAACTCCAGTATCA 152
:::|||||:::|||||:::|||||:::|||||
DB 326 lleaGlYValglYlIleAspGlyArGserleuValThrlyAsnaSerPheArGlyrleu 345
QY 153 ACGACTATTTCAGAAAGTTTTCGACCTGAAACGAGATACAGCTTCTTTGCTGAA 212
346 HlsThrleuIleAsnaeuglySerAlaProgluProAsnaMetThrValleuTrpserGlu 365
QY 213 AATGTTGAAAGAGAGACTTATATTATTAGTGAACCTGACTGCTGAAATGTA 272
|||||:::|||||:::|||||:::|||||
DB 366 Asnaeu---ProgluSerPheLylYelYphecYalaGlumecSerlleleuThrAsper 384
QY 273 TCCGAGTA-----ACGCTTACCTGTTTATGATGAGATATGCTTA--- 317
:::|||||:::|||||:::|||||:::|||||
DB 385 lIleGInlyrGluAsnaAspPleleArGProlelyrGlyAspAspTyAlalIleAla 404
318 -----GAGAGACTAGTAAT 332
QY 405 CysCysValserAlameArGValglYlYsAspMetGInPhePheglYAlaArGysAsn 424
333 ACCGCTGATGCTCCATCAATTAATCTACAGATTGTGAGAGAAAGGAAATATG 392
|||||:::|||||:::|||||:::|||||
DB 425 leuAlalYsCysleuLeuAlalIleAsnGlyglYValAspGluLys----- 440
393 TTAAAGATCATGCTGCGAGAGTTTACA-----CAACAAATTACTTTTAAAGCGCTG 449
441 -----LysglYlIleYValValProAsPlelgluProleThrAspGluValleu 457
QY 450 AATTTACTACCGGGGAAAAAATATCTCTGGAATATTAAGAT---CAGTTATG 506
458 AspTYr-----GluLysVallysgluAsnTyrPheLysValleuGluTywec 473
QY 507 GTTGTACTATGTAAAC 524
DB 474 AlaglyLeuTyryValAsn 479

RESULT 3
US-09-072-433-30
Sequence 30, Application US/09072433
Patent No. 6551814
GENERAL INFORMATION:
APPLICANT: Coschigano, Peter W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,433

FILING DATE: 04-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CHU-03344
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-433-30

Alignment Scores:
Pred. No.: 0.0849 Length: 740
Score: 84.00 Matches: 39
Percent Similarity: 38.6% Conservative: 25
Best Local Similarity: 23.5% Mismatches: 68
Query Match: 8.9% Indels: 34
Gaps: 7

US-10-768-093-4 (1-527) x US-09-072-433-30 (1-740)

```
QY 93 CTTCCTCATTCGGCCACCATGCCAGAACAAATAGGACTAAACCTCCAGTATCA 152
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 326 ILeLaLlYlVAlGlylLeaSPglYArSerLeuValThrLysaenSerPheArTYrLeu 345
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 153 ACGACTATTCACAAAGATTTTTCACCTGACACACAGACAGCTTCCTTGTTGTA 212
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 346 HlThrLeuLeaenLeuGlySerLlaProGluProaenMetThrValLeuTrpSerGlu 365
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 213 AATGTGGAAAGAGAGAGCTTTATTTAGTGTGAACCTTAAGTCTTCGAAATGTA 272
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 366 AenLeu---ProGluSerPheLysPheCysAlaGluMetSerLleuThraSPser 384
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 273 TCCAGGTA-----ACGCTCAACCTGTTATGATGTAAGTATAGGGTAA 317
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 385 ILeGlnTrGluenaSPaSPlleMetArGProLleTYrGlyaSPaSPYAlaIleAla 404
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 318 -----GGACGACTAGTAAT 332
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 405 CysCysValSerAlaMetArGValGlyLysaSPmetGlnPhePheGlyAlaArgCysaSPn 424
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 333 ACCGCTGATGCTCCCAATCAATATCAACAGATTGATGAGAAAGGAAAAAATG 392
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 425 LeuLlaLysCysLeuLeuLeuAlaIleEnGlyGlyValaSPGluLys----- 440
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 393 TTAAAAAGATCATGTGACAGAGTTACACCTAAT---CAACAAATACCTTTAAAGCGCTG 449
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 441 -----LysGlyIleLysValValaProaSPlleGluProLleTrpAspGlyValleu 457
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 450 AATTATCTAGCGGGGAAAAAATATATCTCTGGAATATATTAAGAT---CAGGTATG 506
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 458 AspTYr-----GluLysValLysGluLysValLeuGluTYrMet 473
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 507 GTTGCTACTATGTAAC 524
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 474 AlaeGlyLeuTYrValaSPn 479
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

RESULT 4
US-09-248-796A-14853
Sequence 14853, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 14853
LENGTH: 330
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-14853

Alignment Scores:
Pred. No.: 0.107 Length: 330
Score: 82.00 Matches: 34
Percent Similarity: 37.3% Conservative: 23
Best Local Similarity: 22.2% Mismatches: 44
Query Match: 8.7% Indels: 52
Gaps: 7

US-10-768-093-4 (1-527) x US-09-248-796A-14853 (1-330)

```
QY 150 TCAACGACTATTTCACAAATTTTTCACCTGACACGACAGAAATACAGCCTCTTTGGT 209
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 44 SerSerThrSerSerThrSerThrPheSerProSerPro-----Ser 57
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 210 GAATATGTTGAAGAGAGAGCTTTATTTATTTAGTGAACCTTAAGTCTTCGTAAT 269
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 58 SerSerThrGlySerSerGly-----AspLleProAlaProThraAla 71
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 270 GATATCCAGGTAACGCTTACCTGTTATGATGATGAATATGAGGTAGACGACTAGTA 329
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 72 Ile-----ValTYrSerProTYrAlaAsnSP-----ArgSerCys 83
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 330 AATACCGCTGATGCTCCCAATCAATATCTCAACAGTTGATGAGAAAGGAAAAA 389
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 84 LysSerLlaSPThrLleArGSerAspLle---GlnLeuLlaenSerLysGlyLleLys 102
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 390 ATGTTAAAGATCATGAGTGCAGAG----- 413
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 103 GlnLleArGSerTYrGlyThrAspCysGlySerLeuThrThrValleuGluThrcysArg 122
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 414 -----GTTACACCTTAATCAACAAATACCTTTAAAGCGCTGAATTAATACGCGG 464
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 123 GluLeuGlyLleThrValaSPnGlnGlyValTrpLleSerAlaIleGlyValaSPserIle 142
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 465 GAAAAAATATATCTCT----- 482
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 143 AspAspGlnValaSerAspValIleLysTYrGlyGlnSerAsnGlyTrpAspValaSPaSPn 162
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 483 -----GGAATATATACAGATGAGTTAGTGTGTTAC 515
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 163 LeuLeuThrLleGlyaSPnGlnAlaIleIleIleAlaGlyTYr 175
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

RESULT 5
US-09-134-000C-6467

Sequence 6467, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6467
LENGTH: 768
TYPE: PRT

ORGANISM: Enterococcus faecalis
US-09-134-000C-6467

Alignment Scores:
Pred. No.: 0.215 Length: 768
Score: 81.00 Matches: 43
Percent Similarity: 35.7% Conservative: 31
Best Local Similarity: 20.8% Mismatches: 67
Query Match: 8.6% Indels: 66
DB: 2 Gaps: 9

US-10-768-093-4 (1-527) x US-09-134-000C-6467 (1-768)

```
QY 66 AAGAAACAATGCTTAAATCTTAATTCCTTCCTTCGAGCAGCCATGCCAGAACAGAA 125
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 21 Lysbysanvalargprotyrvalleuvalargpneglyleugluleugln 40
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 126 ATAGCGACTAAATCTCCAGCTATCAAGACTATTTCAAAAGTTT 176
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 41 Argserthrarggluylleuvalathrthrghlmsprothrvalpneglyasna 60
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 177 GCACCTGAACCAAGATCAGCCTCTTGTGTA 212
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 61 Sertryhspotyrlleghlnhraspheserghlnrlnleugluleuthrpro 80
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 213 -AATGTTGGAAGAA- -GGAGCTTATTAATTAGTGAAC 251
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 81 Valalaenserglythrclumetleuargpneleuapalailehisapvalala 100
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 252 TTAAGTCTCTGAAAT- -GATCCAGGTAAAGGTCTACCTGTT 296
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 101 Argserileprogluaspglumetleutrproulsermetproglinleuprothr 120
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 297 TATGATGAAGTTAGGTTAGACGACTAATAATCCGCTGCTCCCAATCAATA 356
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 121 Lysaspglnluilelyleleuvalaleu- -Aspghlntryasphalaival 136
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 357 ATTCACGAG- -ATTGTTGATGAGAAAGGAAAAA- - 389
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 137 Leutyrargagtyrleuvalaleuglutyrglylsargylmetvalserglyle 156
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 390 - - - - -ATCTTAAAGATCATGTCGACAG 413
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 157 Hispheanpheglutyraspglnaleuileghlnleutyraspglnserglu 176
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 414 GTTACACCTAATCAACAATACTTTTAAAGCG- - 446
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 177 ValthrarpCylysglnphelythrlysvailtyrmetlyvalalaalargasnphelu 196
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 447 - - - - -CTGAATTATACTAGCGGGGAAAAAATAATCTCTCGA- - 485
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 197 Argtyrargtrpleuilethrtyrleupheglalasrprovalserclupspargtyr 216
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 486 - - - - -ATATATACGATCAG 500
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 217 PheargvaltyrAspAspGln 223
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::

RESULT 6
US-09-538-092-314
; Sequence 314, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Manfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178, 965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
```

SOFTWARE: Curpatseqformatcer Version 0.9

SEQ ID NO 314
LENGTH: 345
TYPE: PRF
ORGANISM: Saccharomyces cerevisiae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number YGL175C
US-09-538-092-314

Alignment Scores:
Pred. No.: 0.501 Length: 345
Score: 77.00 Matches: 44
Percent Similarity: 40.6% Conservative: 34
Best Local Similarity: 22.9% Mismatches: 82
Query Match: 8.2% Indels: 32
DB: 2 Gaps: 7

US-10-768-093-4 (1-527) x US-09-538-092-314 (1-345)

```
QY 8 ATACGAATGTTATGATATACAAATAAATGATTAACCAATTAATGCTGTATGCA 67
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 25 LeuvalenvalglntryasphvalthrThrleuilealaleysargvalglnaleugln 44
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 68 GAAAAAATGCTTAAATCTTAATCTTCTTCATTCGAGCAACA- - 112
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 45 - - - - -Aaaargasnlyscyvalleugluluproamserlyleu 59
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 113 - - - - -TGCCAGAACAGAAATAGCGACTAAAACTCCAGTATCAAGACTATTC 163
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 60 GlulleuCyhlsghlyasphalpro- -glnlnserSer-GlnhrserAlaGl 78
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 164 AAAAGTTTTCACCTGACCAACCAATACAGCTTCTTGTGTAATGTTGAAA 223
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 78 Yproglgluclnaspsergluasphneilleuthrghlnheaspgluasphlely 98
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 224 GGAAGAGCTTATTAATTAAGTGAAC- - - - -TT 253
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 98 sgluserlaclualvalhstryrargasnluasnlyshthrvalglnleuproleu 118
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 254 AACTGCTCCGAAAATGATATCCAGGTAAAGGTCTACCTGTTATGATGAAGTTATG 313
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 118 lthmepprobsnarghlslyargylsieserclupheserproleuasngl 138
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 314 GTTGAAGCACTAGTAATAACCGCTGATGCTTCCCAATCAATAATCAAGATTGTA 373
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 138 Yleuasnlnleuaserasphleugluaspcysasphthrvalilehlsghlyasphs 158
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 374 TGAAGAAAGGAAAAAATGTTAAAGATCATGTCAGAGGTTACACTAATCAACAAT 433
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 158 nasplysegluasnlyshthrargylsleuuglylsgluu- -gluasnproglu 177
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 434 AACTTTTAAAGCGCTGAATTAATCAAGCGGGGAAAAAATAATCTCTCGAATAT 489
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 177 rThrserproasnleu- -tyrlyasn-Vallyasphasnphleuapheasent 196
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
QY 490 - - - - -ATACGATCAGGTTATGCTGTT 513
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
Db 196 hrasnproleuthrlyasphalatrprileu 206
    |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::

RESULT 7
US-09-252-991A-21015
; Sequence 21015, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

US-10-768-093-4 (1-527) X US-09-543-681A-6955 (1-186)

```

QY      1  TTGACGATTCAGATGTTATGATACAACTAAATATGTTTATGACAAATATTAATGGTGT- 59
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      37  MeThAArgLSeLeuLeThrPrOthNiYsThAArgYrSeAeArAlAnHIGLYasn 56
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      60  -----TATTAAGAAACAACATGGTTTAAATTCATTCCTGGTTCA 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      57  SeArgLGLyALGInGLnTLYeSLyGSLYrLpLLeYrTLPLeYSLLe ----- 70

```


Db 140 LeuAnGlnAlaThrSerGlyAlaAlaValSerHisLeuGlySerAsn 159
OY 369 -----GTTGATGAGAAAGGAAAAAAGTTAAAGATCATGTCGACAG 413
Db 160 AsnAlaValIleProValIleAlaIleGlyAlaSerGlyAlaGlnAAspAsnSerGlnTyr 179
OY 414 GTTACA-----CCTAATCAACAAATACTTTAAAGCGCTGAATTATATAC 461
Db 180 ValAsnValSerAlaGlyAlaAsnAlaAspLeuSerTyrTyrAlaGluTyrTyrAlaThr 199
OY 462 GGGGAAAAAATATCTCTCGAATATATAACATGAGTT 503
Db 200 GlyAlaSer-----ThrAlaGlyThrValThrSerGlnVal 211

RESULT 12
US-09-252-991A-23709
Sequence 23709, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23709
LENGTH: 527
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23709

Alignment Scores:
Pred. No.: 0.827 Length: 527
Score: 76.00 Matches: 27
Percent Similarity: 44.8% Conservative: 16
Best Local Similarity: 28.1% Mismatches: 37
Query Match: 8.1% Indels: 16
Gaps: 5

US-10-768-093-4 (1-527) x US-09-252-991A-23709 (1-527)

OY 78 GGTTAATTCCTTAATTCCTTCATTC-----GGCAGC 110
Db 129 GlyAlaLeuValMetAlaArgPheArgIleLeuGluIleAspProLeuGlyArg 148
OY 111 CATCCAGACAGAAATAGCAGTAAATCCAGTATCAAGCATTTCAAAAAGT 170
Db 149 PheAlaArgValGlnProGlyValAlaArgHisLeuAlaIleSerGlnAlaAlaProHis 168
OY 171 -----TTTTCGACCTGAACCA-----CGAATACAGCTTCTTTGGTGAATGTT 218
Db 169 GlyLeuTyrTyrAlaProAspProSerSerGlnIleAlaCySerIleGlyAlaAsnVal 188
OY 219 GGAAGGAA-----GAGCTTTATATTAGTGAAGTAACTTATCTCCGAAAGTGA 272
Db 189 AlaIleHisAlaGlyGlyValAlaHisCysLeuTyrTyrGlyLeuThrVal--HisAsnLeu 207
OY 273 TCCGAGTTACGCTTACCTGTTATGATGAAGATTTGGGTTAGGA 320
Db 208 LeuGlnValAlaPheIleValThrLeuGluGlyIleArgLeuSerLeuGly 223

RESULT 13
US-09-902-540-11397
Sequence 11397, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Goldman, Gregory J.
APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 11397
LENGTH: 603
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-09-902-540-11397

Alignment Scores:
Pred. No.: 1.03 Length: 603
Score: 75.50 Matches: 30
Percent Similarity: 41.6% Conservative: 17
Best Local Similarity: 26.5% Mismatches: 35
Query Match: 8.0% Indels: 31
Gaps: 6

US-10-768-093-4 (1-527) x US-09-902-540-11397 (1-603)

OY 141 TTCCAGATATCAAG-----ACTATT 161
Db 304 PheProValAspSerAlaAspTyrGluAsnLeuArgAspAlaLeuAlaValLeuValLeu 323
OY 162 TCAAAAAGTTTTCGACCTGAACCAAGATACAGCTTCT-----TTGGTGAAAT 215
Db 324 AsnAspSerAlaPheThrTyrGluProGluSerSerThrAlaLeuGlyPheGlyPheArg 343
OY 216 GTTGAAAGGAGAGGCTTATATTATTAGTGAAGTAACTTATCTCGTGAAGATGATCC 275
Db 344 CysGlyTyrLeuGlyLeuLeuHisMetGluIle-----ValGlnGluArgLeu--- 359
OY 276 CAGGTACGGTCTACCTGTTATGATGAAGATTAAGGTTAGGACATGATTAATACC 335
Db 360 -----GluArgGluTyrAsnLeu--AsnLeuIleThrThr 370
OY 336 GGTGATGCTTCCCAATGATATCTACAGATTTGTGATGAGAAAGGAAAAATGTTA 395
Db 371 Ala-----ProSerValValTyrArgIleThrThrSerTyrGlyGluThrLeu 387
OY 396 AAAGATCATGTCAGAGGTTTACCACTTATCAACAATA 434
Db 388 ValAspAsnProAlaTyrLeuProProValGlnHisIle 400

RESULT 14
US-09-541-782-2
Sequence 2, Application US/09541782
Patent No. 6284480
GENERAL INFORMATION:
APPLICANT: Nislow, Corey
APPLICANT: Sakowicz, Roman
APPLICANT: Berand, Christophe
TITLE OF INVENTION: Antifungal Assay
FILE REFERENCE: 1015
CURRENT APPLICATION NUMBER: US/09/541,782
CURRENT FILING DATE: 2000-04-03
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1184
TYPE: PRT
ORGANISM: A. nidulans
US-09-541-782-2

Alignment Scores:
Pred. No.: 1.41 Length: 1184
Score: 75.50 Matches: 41
Percent Similarity: 38.6% Conservative: 23

Best Local Similarity: 24.7% Mismatches: 57
 Query Match: 8.0% Indels: 45
 DB: 2 Gaps: 8

US-10-768-093-4 (1-527) x US-09-541-782-2 (1-1184)

```

Qy 72 ACAATGTTTAATTCCTTAATTCCTTCATTCGGACGACCATGCCAGAACAGAAATACCG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ThrAlaGlyIleLeuLeu-----GInGInGlySerHisLeuValAlaIaIa 282
Qy 132 ACTAA-----AACTCCAGATACAGACTATT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ThrIysCysAsnApeLeuSerSerArgSerHisThrValPheThrIleThrValAsnIle 302
Qy 162 TCAGAAAGTTTTTTTGACCTGAAACACAGATACAGCT----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LysArgThrThrGInSerGlyGInGluIuTyValCysProGlyLysLeuAsnLeuValasp 322
Qy 201 --TCCTTGGTGAATGTGTGGAAGAGAGAGCTTTATTATTATTAGTGGAATTAACT 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 LeuAlaGlySerGluAsnIleGlyArgSerGlyAla----- 334
Qy 258 GTTCCTGAATGTATCCAGGTACAGGTCTACCTGTTTATGATGAAGATTATGGGTTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 -----GluAsnLysArgAlaThrGluAlaGlyLeuIleAsnLysSerLeuThrLeu 352
Qy 318 GGACGACTAGTAATACCC--GCTGATGCTTCCCAATCAATCAATCAATCAATGTTGAT 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 GlyArgValIleAsnAlaLeuValAspLysSerGlnHisIleProTyArg-----Glu 370
Qy 375 GAGAAAGGAAAAAATGTTAAAGATCAT----- 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 SerIysLeuThrArgLeuGlnAspSerLeuGlyArgThrLysThrCysIleIle 390
Qy 408 GCAGAGTTTACACT-----AATCAACAATAATCTTTAAAGCGCTGAATTACT 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AlaThrMetSerProAlaArgSerAsnLeuGluIuThrIleSerThrLeuAspTyAla 410
Qy 459 AGCGGGGAAAAAATA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 PheArgAlaLysAsnIle 416
  
```

RESULT 15
 US-09-723-820-2

; Sequence 2, Application US/09723820

; Patent No. 6468760

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: 09/541,782

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: A. nidulans

; US-09-723-820-2

Alignment Scores:

Pred. No.: 1.41 Length: 1184
 Score: 75.50 Matches: 41
 Percent Similarity: 38.6% Conservative: 23
 Best Local Similarity: 24.7% Mismatches: 57
 Query Match: 8.0% Indels: 45
 DB: 2 Gaps: 8

US-10-768-093-4 (1-527) x US-09-723-820-2 (1-1184)

```

Qy 72 ACAATGTTTAATTCCTTAATTCCTTCATTCGGACGACCATGCCAGAACAGAAATACCG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ThrAlaGlyIleLeuLeu-----GInGInGlySerHisLeuValAlaIaIa 282
Qy 132 ACTAA-----AACTCCAGATACAGACTATT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ThrIysCysAsnApeLeuSerSerArgSerHisThrValPheThrIleThrValAsnIle 302
Qy 162 TCAGAAAGTTTTTTTGACCTGAAACACAGATACAGCT----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LysArgThrThrGInSerGlyGInGluIuTyValCysProGlyLysLeuAsnLeuValasp 322
Qy 201 --TCCTTGGTGAATGTGTGGAAGAGAGAGCTTTATTATTATTAGTGGAATTAACT 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 LeuAlaGlySerGluAsnIleGlyArgSerGlyAla----- 334
Qy 258 GTTCCTGAATGTATCCAGGTACAGGTCTACCTGTTTATGATGAAGATTATGGGTTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 -----GluAsnLysArgAlaThrGluAlaGlyLeuIleAsnLysSerLeuThrLeu 352
Qy 318 GGACGACTAGTAATACCC--GCTGATGCTTCCCAATCAATCAATCAATCAATGTTGAT 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 GlyArgValIleAsnAlaLeuValAspLysSerGlnHisIleProTyArg-----Glu 370
Qy 375 GAGAAAGGAAAAAATGTTAAAGATCAT----- 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 SerIysLeuThrArgLeuGlnAspSerLeuGlyArgThrLysThrCysIleIle 390
Qy 408 GCAGAGTTTACACT-----AATCAACAATAATCTTTAAAGCGCTGAATTACT 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AlaThrMetSerProAlaArgSerAsnLeuGluIuThrIleSerThrLeuAspTyAla 410
Qy 459 AGCGGGGAAAAAATA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 PheArgAlaLysAsnIle 416
  
```

Search completed: April 10, 2006, 10:08:36
 Job time : 38.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 10, 2006, 10:07:45 ; Search time 25.1 Seconds
(without alignments)
1754.550 Million cell updates/sec

Title: US-10-768-093-4
Sequence: 1 TTGCACATTAAGATGTTA.....TTGTACTATGTAAGTAA 527

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-Q=/abse/ABSSWB_topool/US10768093/runat_10042006_090309_4677/app_query.fasta_1
-DB=Published_Applications_AA_Main -QFWT=fastan -SUFFIX=n2p.rapbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pct -NORM=ext
-HEAFSIZE=500 -MITLEN=0 -MAXLEN=2000000000 -HOST=aab6602p
-USRR=US10768093 @CN 1.1 307 @runat_10042006_090309_4677 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	785	83.6	154	4	US-10-768-093-5
2	785	83.6	154	4	US-10-754-641-5
3	699	74.4	136	4	US-10-768-093-9
4	699	74.4	136	4	US-10-754-641-9
5	118	12.6	167	4	US-10-768-093-6
6	118	12.6	167	4	US-10-754-641-6
7	113.5	12.1	146	4	US-10-768-093-10
8	87.5	9.3	911	4	US-10-197-666A-100
9	85	9.1	541	4	US-10-282-122A-77158
10	84	8.9	439	4	US-10-365-493-2489
11	84	8.9	740	4	US-10-267-989-25

12	84	8.9	740	4	US-10-357-567-25	Sequence 25, Appl
13	83.5	8.9	157	4	US-10-424-599-190153	Sequence 190153,
14	83	8.8	249	4	US-10-424-599-250712	Sequence 250712,
15	83	8.8	750	5	US-10-501-282-3988	Sequence 3988, Ap
16	82	8.8	2094	4	US-10-437-963-161607	Sequence 161607,
17	79.5	8.5	429	4	US-10-359-493-17433	Sequence 17433, A
18	79.5	8.5	951	4	US-10-197-666A-94	Sequence 94, Appl
19	79.5	8.5	1171	4	US-10-197-666A-96	Sequence 96, Appl
20	79.5	8.5	1171	4	US-10-311-034-8	Sequence 8, Appl
21	78.5	8.4	1198	4	US-10-001-215-2	Sequence 2, Appl
22	78.5	8.4	1198	4	US-10-197-666A-98	Sequence 98, Appl
23	77.5	8.3	357	4	US-10-282-122A-65978	Sequence 65978, A
24	77	8.2	458	4	US-10-425-115-223762	Sequence 223762,
25	77	8.2	471	3	US-09-815-242-11936	Sequence 11936, A
26	76.5	8.1	330	4	US-10-815-242-11936	Sequence 7096, Ap
27	76.5	8.1	442	4	US-10-452-024-76	Sequence 76, Appl
28	76.5	8.1	1076	4	US-10-724-972A-5848	Sequence 5848, Ap
29	76	8.1	348	4	US-10-437-963-136628	Sequence 136628,
30	76	8.1	499	4	US-10-389-647-664	Sequence 664, App
31	76	8.1	557	4	US-10-369-493-21384	Sequence 21384, A
32	76	8.1	2358	4	US-10-282-122A-68607	Sequence 68607, A
33	75.5	8.0	246	4	US-10-282-122A-67315	Sequence 67315, A
34	75.5	8.0	1218	5	US-10-470-048B-444	Sequence 444, App
35	75.5	8.0	1224	4	US-10-282-122A-70506	Sequence 70506, A
36	75	8.0	640	4	US-10-282-122A-44691	Sequence 44691, A
37	75	8.0	736	3	US-09-815-242-11293	Sequence 11293, A
38	75	8.0	736	4	US-10-282-122A-58624	Sequence 58624, A
39	74.5	7.9	401	4	US-10-282-122A-54471	Sequence 54471, A
40	74.5	7.9	430	4	US-10-282-122A-60954	Sequence 60954, A
41	74.5	7.9	497	4	US-10-724-972A-6030	Sequence 6030, Ap
42	74.5	7.9	499	4	US-10-369-493-23554	Sequence 23554, A
43	74.5	7.9	581	5	US-10-450-763-48573	Sequence 48573, A
44	74.5	7.9	2849	4	US-10-437-963-151131	Sequence 151131,
45	74	7.9	1666	6	US-11-097-143-42954	Sequence 42954, A

ALIGNMENTS

RESULT 1
US-10-768-093-5
Sequence 5, Application US/10768093
Publication NO. US20040156828A1
GENERAL INFORMATION:
APPLICANT: Wolf, Marcia K
Cassels, Fred J
Boedecker, Edgar C
TITLE OF INVENTION: Transformed Bacteria Producing GS6
Antigens as Vaccines
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hendricks and Assoc
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10768, 093
FILING DATE: 02-Feb-2004
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/479, 877B
FILING DATE: 10-Jan-2000
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna M
REGISTRATION NUMBER: 32,535
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/425-8405

TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 154 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-10-768-093-5

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.09e-82	785.00	100.0%	100.0%	83.6%	154	154	0	0	0	0

US-10-768-093-4 (1-527) x US-10-768-093-5 (1-154)

```

OY 63 ATGAAGAAAACAAATGGTTTAAATCTTCTGCTTCAATTCGGCAGCCATGCCAGACA 122
DB 1 MetLysLysThrIleGlyLeuIleLeuIleLeuIleSerPheGlySerHisAlaArgThr 20
OY 123 GAAATAGGACTAAACCTCCAGTATCAACGACTATTCCTTCAAAAAGTTTGTTCACCT 182
DB 21 GluIleAlaThrLysAsnProValSerThrThrIleSerLysSerPhePheAlaPro 40
OY 183 GAACCAAGAAATACAGCCCTCTTTGGTGAATAATGTTGAAAGAGAGACTTATATT 242
DB 21 GluIleAlaThrLysAsnProValSerThrThrIleSerLysSerPhePheAlaPro 40
OY 243 AGTGTGAATTAATCTGTTCTGAAAATGTATCCCAAGTAAACGCTTACCTGTTATGAT 302
DB 41 GluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeuLeuPhe 60
OY 41 GluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeuLeuPhe 60
DB 61 SerValAsnLeuThrValProGluAsnValSerGluValThrValTyProValTyAsp 80
OY 303 GAAGATTATGGGTTAGAGCAGCTAGTAAATACCGCTGATGCTTCCCAATCAATATCTAC 362
DB 81 GluMetPylrGlyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIleIleTy 100
OY 363 CAGATTGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGCGAGGTTACACCT 422
DB 101 GlnIleValaArgGluLysGlyLysLysMetLeuLysAspHisGlyAlaGluValThrPro 120
OY 423 AATCAACAAATTAATCTTTTAAAGCGCTGAATTAATCTAGCGGGGAAAAAATATCTCCT 482
DB 121 AsnGlnGlnIleThrPheLysAlaLeuAsnTyThrSerGlyGluLysLysIleSerPro 140
OY 483 GGAATATATAAGATCAGATTATGTTGTTACTATGTTAAAC 524
DB 141 GlyIleTyraAsnAspGlnValMetValGlyTyTyTyValaAsn 154

```

RESULT 2

US-10-754-641-5

Sequence 5, Application US/10754641

Publication No. US20050025787A1

GENERAL INFORMATION:

APPLICANT: U.S. Army Medical Research and Materiel Command

APPLICANT: Wolf, Marcia K.

APPLICANT: Casseels, Frederick J.

APPLICANT: Boedeker, Edgar C.

TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES

FILE REFERENCE: 01047.033.3

CURRENT FILING DATE: 2004-01-12

PRIOR APPLICATION NUMBER: 09/479,877

PRIOR FILING DATE: 2000-01-10

PRIOR APPLICATION NUMBER: 08/788,145

PRIOR FILING DATE: 1997-01-24

PRIOR APPLICATION NUMBER: 08/243,482

PRIOR FILING DATE: 1994-05-13

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.2

SEQ ID NO 5

LENGTH: 154

TYPE: PRT

ORGANISM: Escherichia coli

US-10-754-641-5

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1.09e-82	785.00	100.0%	100.0%	83.6%	154	154	0	0	0	0

US-10-768-093-4 (1-527) x US-10-754-641-5 (1-154)

```

OY 63 ATGAAGAAAACAAATGGTTTAAATCTTCTGCTTCAATTCGGCAGCCATGCCAGACA 122
DB 1 MetLysLysThrIleGlyLeuIleLeuIleLeuIleSerPheGlySerHisAlaArgThr 20
OY 123 GAAATAGGACTAAACCTCCAGTATCAACGACTATTCCTTCAAAAAGTTTGTTCACCT 182
DB 21 GluIleAlaThrLysAsnProValSerThrThrIleSerLysSerPhePheAlaPro 40
OY 183 GAACCAAGAAATACAGCCCTCTTTGGTGAATAATGTTGAAAGAGAGACTTATATT 242
DB 21 GluIleAlaThrLysAsnProValSerThrThrIleSerLysSerPhePheAlaPro 40
OY 243 AGTGTGAATTAATCTGTTCTGAAAATGTATCCCAAGTAAACGCTTACCTGTTATGAT 302
DB 41 GluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeuLeuPhe 60
OY 41 GluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeuLeuPhe 60
DB 61 SerValAsnLeuThrValProGluAsnValSerGluValThrValTyProValTyAsp 80
OY 303 GAAGATTATGGGTTAGAGCAGCTAGTAAATACCGCTGATGCTTCCCAATCAATATCTAC 362
DB 81 GluMetPylrGlyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIleIleTy 100
OY 363 CAGATTGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGCGAGGTTACACCT 422
DB 101 GlnIleValaArgGluLysGlyLysLysMetLeuLysAspHisGlyAlaGluValThrPro 120
OY 423 AATCAACAAATTAATCTTTTAAAGCGCTGAATTAATCTAGCGGGGAAAAAATATCTCCT 482
DB 121 AsnGlnGlnIleThrPheLysAlaLeuAsnTyThrSerGlyGluLysLysIleSerPro 140
OY 483 GGAATATATAAGATCAGATTATGTTGTTACTATGTTAAAC 524
DB 141 GlyIleTyraAsnAspGlnValMetValGlyTyTyTyValaAsn 154

```

RESULT 3

US-10-768-093-9

Sequence 9, Application US/10768093

Publication No. US20040156829A1

GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K.

APPLICANT: Casseels, Fred J.

APPLICANT: Boedeker, Edgar C.

TITLE OF INVENTION: Transformed Bacteria Producing CS6

Antigens as Vaccines

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESS: Hendricks and Assoc

STREET: P.O. Box 2509

CITY: Fairfax

STATE: VA

COUNTRY: US

ZIP: 22031

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/768,093
FILING DATE: 02-Feb-2004
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/479,877B
FILING DATE: 10-Jan-2000
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna M
REGISTRATION NUMBER: 32,535
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703/425-8405
TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-768-093-9

Alignment Scores:
Pred. No.: 1,23e-72 Length: 136
Score: 699.00 Matches: 136
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 74.4% Indels: 0
DB: Gaps: 0

US-10-768-093-4 (1-527) x US-10-768-093-9 (1-136)

QY 117 AGAACAAGTAAGGACTAAAGCTCCAGTATCAACGACTATTTCAGAAAGTTT 176
Db 1 ArgThcIuIleAthrLysAsnPhcProValSerThrThrIleSerLysSerPhe 20

QY 177 GCACCTGAACCAAGATACAGCCTCTTTGGTGAATAATGTGGAAGAGAGCTTTA 236
Db 21 AlaProGluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeu 40

QY 237 TTAATTAGTGAACCTTAAGTCTCTGAAATGTATCCAGTAAGGCTTACCTGTT 296
Db 41 LeuPheSerValAsnLeuThrValProGluAsnValSerGlnValThrValTyProVal 60

QY 297 TATGATGAAGATTATGGGTTAGAGCACTAGTAATAATCCGCTGATGCTTCCCAATCAATA 356
Db 61 TyAspGluAspTyrGlyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIle 80

QY 357 ATTCACGAGATTGTGTGTAAGAAAGGAAAAAATGTTAAAGATCATGTGCGAGAGTT 416
Db 81 IleTyrGlnIleValAspGluLysGlyLysValSerLeuLysAspHisGlyAlaGluVal 100

QY 417 ACACTATCAACAATAACTTTTAAAGCGCTGATTATACGCGGGAATAATAATA 476
Db 101 ThrProAsnGlnGlnIleThrPheLysAlaLeuAsnTyrThrSerGlyGluLysIle 120

QY 477 TCTCCGGAATATATTAACGATCAGGTTATGGTGTACTATGTAAC 524
Db 121 SerProGlyIleTyrAsnAspGlnValMetValGlyTyrValAsn 136

RESULT 4
US-10-754-641-9
Sequence 9, Application US/10754641
Publication No. US20050025787A1
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research and Materiel Command
APPLICANT: Wolf, Marcia K.
APPLICANT: Casseels, Frederick J.
APPLICANT: Boedeker, Edgar C.

TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES
FILE REFERENCE: 034047.033.3
CURRENT APPLICATION NUMBER: US/10/754,641
CURRENT FILING DATE: 2004-01-12
PRIOR APPLICATION NUMBER: 09/479,877
PRIOR FILING DATE: 2000-01-10
PRIOR APPLICATION NUMBER: 08/788,145
PRIOR FILING DATE: 1997-01-24
PRIOR APPLICATION NUMBER: 08/243,482
PRIOR FILING DATE: 1994-05-13
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin version 3.2
SEQ ID NO 9
LENGTH: 136
TYPE: PRT
ORGANISM: Escherichia coli
US-10-754-641-9

Alignment Scores:
Pred. No.: 1,23e-72 Length: 136
Score: 699.00 Matches: 136
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 74.4% Indels: 0
DB: Gaps: 0

US-10-768-093-4 (1-527) x US-10-754-641-9 (1-136)

QY 117 AGAACAAGTAAGGACTAAAGCTCCAGTATCAACGACTATTTCAGAAAGTTT 176
Db 1 ArgThcIuIleAthrLysAsnPhcProValSerThrThrIleSerLysSerPhe 20

QY 177 GCACCTGAACCAAGATACAGCCTCTTTGGTGAATAATGTGGAAGAGAGCTTTA 236
Db 21 AlaProGluProArgIleGlnProSerPheGlyGluAsnValGlyLysGluGlyAlaLeu 40

QY 237 TTAATTAGTGAACCTTAAGTCTCTGAAATGTATCCAGTAAGGCTTACCTGTT 296
Db 41 LeuPheSerValAsnLeuThrValProGluAsnValSerGlnValThrValTyProVal 60

QY 297 TATGATGAAGATTATGGGTTAGAGCACTAGTAATAATCCGCTGATGCTTCCCAATCAATA 356
Db 61 TyAspGluAspTyrGlyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIle 80

QY 357 ATTCACGAGATTGTGTGTAAGAAAGGAAAAAATGTTAAAGATCATGTGCGAGAGTT 416
Db 81 IleTyrGlnIleValAspGluLysGlyLysValSerLeuLysAspHisGlyAlaGluVal 100

QY 417 ACACTATCAACAATAACTTTTAAAGCGCTGATTATACGCGGGAATAATAATA 476
Db 101 ThrProAsnGlnGlnIleThrPheLysAlaLeuAsnTyrThrSerGlyGluLysIle 120

QY 477 TCTCCGGAATATATTAACGATCAGGTTATGGTGTACTATGTAAC 524
Db 121 SerProGlyIleTyrAsnAspGlnValMetValGlyTyrValAsn 136

RESULT 5
US-10-768-093-6
Sequence 6, Application US/10768093
Publication No. US20060156829A1
GENERAL INFORMATION:
APPLICANT: Wolf, Marcia K
APPLICANT: Casseels, Fred J
APPLICANT: Boedeker, Edgar C
TITLE OF INVENTION: Transformed Bacteria Producing GS6
Antigens as Vaccines
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: Hendricks and Assoc
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US

471 AAATATCTCTGGAATATATACGATCAGTTAGTTGTTACTATGTAAC 524
150 GlyTyrSerSerGlyThrTyrAlaGlyAsnLeuThrValSerPheTyrSerAsn 167

RESULT 7

US-10-768-093-10
Sequence 10, Application US/10768093
Publication No. US20040156829A1
GENERAL INFORMATION:
APPLICANT: Wolf, Marcia K
Casele, Fred J
Boedeker, Edgar C
TITLE OF INVENTION: Transformed Bacteria Producing GSE
Antigens as Vaccines
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hendricks and Assoc
STREET: P.O. Box 2509
CITY: Fairfax
STATE: VA
COUNTRY: US
ZIP: 22031

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/768,093
FILING DATE: 02-Feb-2004
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/479,877B
FILING DATE: 10-Jan-2000
ATTORNEY/AGENT INFORMATION:
NAME: Hendricks, Glenna M
REGISTRATION NUMBER: 32,535
TELEPHONE: 703/425-8405
TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-768-093-10
Alignment Scores:
Pred. No.: 0.000436 Length: 146
Score: 113.50 Matches: 43
Percent Similarity: 43.0% Conservative: 22
Best Local Similarity: 28.5% Mismatches: 55
Query Match: 12.1% Indels: 31
DB: Gaps: 7

US-10-768-093-4 (1-527) x US-10-768-093-10 (1-146)
135 AAAAATCTCCAGATCAAGACTATTTCAAAAGTTTTCACCTGAA----- 185
6 LysSerLeuAlaValAsnValAsnValLeuGluGlnAsnPheLeuProAspLeuSer 24
186 ---CCACGAATACAGCCCTTTCTTTGGTGAATAATGTTGAAAGAGAGCTTATATT 242
25 AlaValArgLeuLeuProValAsnTyrAspSerAspProValLeuAspSerGlnLeuTyr 44
243 AGTGTGAATTAATCTTCTCGAAATGTAATCCCGATACGCTTACCC----- 293
45 ThrValGluMetThrIleProAlaGlyValSerAlaValLysIleValIleProThrAspSer 64

294 GTTATGATGAAGATTATGCGTTAGACAGACTGATAATACCGCTGATGCTCCCATCA 353
65 LeuThrSerSerGlyGlnGlnIleGlyLeuValAsnValAsnAsnProAspGlnAsn 84
354 ATATATCCAGATGTTGTATGATGAGAAAGGAAAAAATGTTAAAGATCGTGCA--- 410
85 MetAsnTyrTyrIle-----ArgLysAspSerGlyAlaGly 96
411 -----GAGTTACACCTTAATCAACAATTAAGTTT 440
97 AsnPheMetAlaGlyGlnLysGlySerPheProValLysGluMetThrSerTyr-ThrPhe 116
441 AAAGCGCTGAATTAATCTACCGGGGAA-----AAAAATATCTCTCGAATATAT 491
117 SerAlaIle---TyrThrGlyGlyGlyTyrProAsnSerGlyTyrSerSerGlyThrTyr 135
492 AACGATCAGTTATGTTGTTGTTACTATGTAAC 524
136 AlaGlyAsnLeuThrValSerPheTyrSerAsn 146

RESULT 8

US-10-197-666A-100
Sequence 100, Application US/10197666A
Publication No. US20030092037A1
GENERAL INFORMATION:
APPLICANT: ASahi KASEI KABUSIKI KAISYA
TITLE OF INVENTION: Bkl1 phosphorylation related gene
FILE REFERENCE: PH-1548US
CURRENT APPLICATION NUMBER: US/10/197,666A
CURRENT FILING DATE: 2002-11-18
PRIOR APPLICATION NUMBER: JP 2001-218204
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: JP 2001-263450
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: JP 2002-012176
PRIOR FILING DATE: 2002-01-21
PRIOR APPLICATION NUMBER: US 60/305,884
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/316,304
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: US 60/350,027
PRIOR FILING DATE: 2002-01-23
NUMBER OF SEQ ID NOS: 156
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 100
LENGTH: 911
TYPE: PRT
ORGANISM: Homo sapiens
US-10-197-666A-100

Alignment Scores:
Pred. No.: 0.786 Length: 911
Score: 87.50 Matches: 40
Percent Similarity: 43.3% Conservative: 25
Best Local Similarity: 26.7% Mismatches: 62
Query Match: 9.3% Indels: 23
DB: Gaps: 8

US-10-768-093-4 (1-527) x US-10-197-666A-100 (1-911)
93 CTTCGCTTCATTCGGACCCATGCGACAGAAATACGACTAATAAATCTCCAGATCA 152
34 MetThrGlyTyrGlySerHisSerLysValTyrSerGlnSerLysAsnIleProLeuSer 53
153 ---ACGACTATTTCAAAAGTTTTCGACCTGAAACGACGAATACAG--- 197
54 GlnProAlaThrThrThrValSerThrSerLeuProValProAsnProSerLeuProTyr 73
198 -----CCTTCCTTTGTTGTAATAATGTTGAAAGAGAGCTTATTA 239
74 GluGlnThrIleValPheProGlySerThrGlyHisIleValIleThrSerAlaSerSer 93

```

QY      240  TTAGGTGTAAC-----TTAAGTGTCTCGAAATGTA-----TCCAGATA 281
           |||||
Db      94  ThSerValThrGlyValLeuGlyGlyProHisLeuLeuMetArgGlySerThrVal 1133
QY      282  ACCGCTACCCCTGTTATGATGACATTTATGGCTTAGGACGCACTAGTAATAACCGCTGAT 341
           :::::
Db      114  SerLeuLeuAspThrTyr---GlnTyrCysGlyLeuTyrAspTyrSerGlnGluLeu 1322
QY      342  GCTTCCCAATCATTAAATACCAAGATGTTGTGATGAGAAAGGAAAAAATGTTAAAGAT 401
           :::::
Db      133  AsnThrSerSerVal-----GlnLeuLeuGlnGlnHis---ProPromeIleGlnAsn 149
QY      402  CATGTGTCAGAGGTTACACCTTAATCAACAATAACTTTAAAGCCGCTGAATTAATACGAC 461
           :::::
Db      150  AsnAlaSerGlyAlaThrValAlaThrAlaThrThrPheThrAlaThrSerTyrAsnSer 169
QY      462  GGGGAAAAAAAATATATCTCTCGTAATATAT 491
           |||||
Db      170  GlySerAsn-----SerGlnGlyAspTyr 177

```

```

RESULT 9
US-10-282-122A-77158
/ Sequence 77158, Application US/10282122A
/ Publication No. US2004002912SA1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zyekind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trewick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: EPIIPIA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/268,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77158
/ LENGTH: 541
/ TYPE: PRT
/ ORGANISM: Vibrio cholerae
US-10-282-122A-77158

Alignment Scores:      1.34      Length:      541
Pred. No.:             85.00      Matches:
Score:                  51

```

```
Percent Similarity: 36.2% Conservative: 17
Best Local Similarity: 27.1% Mismatches: 62
Query Matchn: 9.1% Indels: 58
DB: 4 Gaps: 10
```

US-10-768-093-4 (1-527) x US-10-282-122A-77158 (1-541

[illegible]

```

RESULT 10
US-10-369-493-2489
? Sequence 2489, Application US/10369493
? Publication NO. US20030233675A1
? GENERAL INFORMATION:
? APPLICANT: Gao, Yongwei
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
? APPLICANT: Goldman, Barry S.
? APPLICANT: Chen, Xianfeng
? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
? FILE REFERENCE: 38-10(52052)B
? CURRENT APPLICATION NUMBER: US/10/369,493
? CURRENT FILING DATE: 2003-02-28
? PRIOR APPLICATION NUMBER: US 60/360,039
? PRIOR FILING DATE: 2002-02-21
? NUMBER OF SEQ ID NOS: 47374
? SEQ ID NO 2489
? LENGTH: 439
? TYPE: PRT
? ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2489

Alignment Scores: 1.66 Length: 439
Pred. No.:

```

Score: 84.00 Matches: 43
Percent Similarity: 40.4% Conservative: 24
Best Local Similarity: 25.9% Mismatches: 65
Query Match: 8.9% Indels: 34
Gaps: 9

US-10-768-093-4 (1-527) x US-10-369-493-2489 (1-439)

```
OY 54 TGTGTTATATGAGAAACAATTGTT-----TTAATTCTTAATCTTGCTTCA 101
Db 131 TtrpYstrYrIleAlaGluAenPheGlyThrIysGlyProTyrValLeuProValProser 150
OY 102 TTC-----GGCAGCCATCCCGAGACAGAAATAGCAGCTTAAACTCCCA 146
Db 151 PheAenValLeuAenGlySerHisAlaGlyGlyAspLeuAlaPheGlnGluPheMet 170
OY 147 GTATCAACGACTATTTCAAAAGTTT----- 173
Db 171 IleuProThrGlyAlaProSerPheSerGluAlaMetArgTgPlyAlaGluThrTyr 190
OY 174 -----TTGCACCTGAACACGAATACAGCCTCTTTTGTTGAATGTTGA 221
Db 191 HisthrleuYsSerIleAlaIleYsArgTyrGlySerSerAlaGly--AsnValGly 209
OY 222 AAGGAAGAGCTTATTATTATTTAGTGTGAATTAAGTCTTGAAATGTATCCAGGTA 281
Db 210 AspGluGlyGlyIle-----AlaProAspLeuGlnThrProGlnGluAlaLeuAspLeu 227
OY 282 ACGGTTCACCTGTTTATGATGAGATTAATGGGTTAGACACGACTGTAATACCCCTGAT 341
Db 228 IleValGluAlaIleAenIleYsAlaGlyIleGlu--GlyIleIleYsIleGlyLeuAsp 246
OY 342 GCTTCCCATCATATATATATACAGATGTTGTATGAGAAAGGAAAAATGTAAAAAGAT 401
Db 247 ValAlaSerSerGluPheTyr-----ValAsp-----GlyIysTyrAspLeuAspIle 262
OY 402 CATGGTCAGAGGTTATACCTTAATCAACAAATTAACCTTTAAAGCCGTGAAT-----TAT 455
Db 263 LysAlaIleYsProIleProGluAenIleYsLeuThrTyrGlnGlnLeuThrAspLeuTyr 282
OY 456 ACTAGCCGGGAAAAAAA 473
Db 283 ValGluLeuSerIleYs 288
```

RESULT 11

US-10-267-989-25

Sequence 25, Application US/10267989
Publication No. US20030199035A1

GENERAL INFORMATION:

APPLICANT: ARNAU, Jose
APPLICANT: VRANU, Astrid
APPLICANT: ISRAELSEN, Hans
APPLICANT: JOERGENSEN, Flemming
APPLICANT: MADSEN, Soeren
TITLE OF INVENTION: METABOLICALLY ENGINEERED LACTIC ACID BACTERIA AND
TITLE OF INVENTION: MEANS FOR PROVIDING
FILE REFERENCE: ARNAU-1A
CURRENT APPLICATION NUMBER: US/10/267,989
CURRENT FILING DATE: 2002-10-10
PRIOR APPLICATION NUMBER: US/08/981,097
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: PCT/DK97/00336
PRIOR FILING DATE: 1997-08-20
PRIOR APPLICATION NUMBER: 08/701,458
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn version 3.0
SEQ ID NO 25
LENGTH: 740
TYPE: PRT
ORGANISM: Clostridium pasteurianum
US-10-267-989-25

Alignment Scores:
Pred. No.: 1.91 Length: 740
Score: 84.00 Matches: 39
Percent Similarity: 38.6% Conservative: 25
Best Local Similarity: 23.5% Mismatches: 68
Query Match: 8.9% Indels: 34
Gaps: 7

US-10-768-093-4 (1-527) x US-10-267-989-25 (1-740)

```
OY 93 CTTCGCTTCATTCGGCAGCCATGCGAGAAAGAAATAGGACTTAAAACTTCCAGTATCA 152
Db 326 IleAlaGlyValGlyIleAspGlyArgSerLeuValThrIysAenSerPheArgTyrLeu 345
OY 153 ACGACTATTTCAAAAGTTTTCCTGACCTGACACGAGAAATACAGCCTTTTTCGTGAA 212
Db 346 HisthrleuIleAenLeuGlySerAlaProGluProAsnMetThrValLeuTyrSerGlu 365
OY 213 AATGTTGAAAGGAGAGCTTTATTATTATGTTGAATTAAGTCTTCTGAAAAATGTA 272
Db 366 AsnLeu--ProGluSerPheIleYsPheCysAlaGluMetSerIleLeuThrAspSer 384
OY 273 TCCAGGTA-----ACGCTACCCCTGTTTATGATGTAAGATTATGGGTTA-- 317
Db 385 IleGlnTyrGluAenAspAspIleMetArgProIleTyrGlyAspAspTyrAlaIleAla 404
OY 318 -----GGACGACTAGTAAT 332
Db 405 CysCysValSerAlaMetArgValGlyIleYsAspMetGlnPhePheGlyAlaArgCysAsn 424
OY 333 ACCGCTGATGCTTCCCATCATATATATACAGATGTTGTATGAGAAAGGAAAAAATG 392
Db 425 LeuAlaIleYsCysLeuLeuAlaIleAsnGlyIleValAlaAspGluIys----- 440
OY 393 TTAATAATCATGTGTGAGAGGTTTACACTTAAT-----CAACAATATCTTTAAAGCGCTG 449
Db 441 -----LysGlyIleYsValValProAspIleGluProIleThrAspGluValLeu. 457
OY 450 AATTATACTAGCGGGGAAAAAATATATCTCTGAAATATATTAAGAT--CAGGTTATG 506
Db 458 AspTyr-----GluYsValIleGluAsnTyrPheYsValLeuGluTyrMet 473
OY 507 GTTGTTACTATGTAAAC 524
Db 474 AlaGlyLeuTyrValAen 479
```

RESULT 12

US-10-357-567-25

Sequence 25, Application US/10357567
Publication No. US20040038382A1

GENERAL INFORMATION:

APPLICANT: Coschigano, Peter
TITLE OF INVENTION: Compositions and Methods for Bioremediation
FILE REFERENCE: OHU-07748
CURRENT APPLICATION NUMBER: US/10/357,567
CURRENT FILING DATE: 2003-02-04
PRIOR APPLICATION NUMBER: 09/072,433
PRIOR FILING DATE: 1998-05-04
PRIOR APPLICATION NUMBER: 60/046,845
PRIOR FILING DATE: 1997-05-05
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 25
LENGTH: 740
TYPE: PRT
ORGANISM: Clostridium pasteurianum
US-10-357-567-25

Alignment Scores:
Pred. No.: 1.91 Length: 740
Score: 84.00 Matches: 39
Percent Similarity: 38.6% Conservative: 25

Best Local Similarity: 23.5%
Query Match: 8.9%
DB: 4
Gaps: 7

US-10-768-093-4 (1-527) x US-10-357-567-25 (1-740)

QY 93 CTGCTTCATTCGGCAGCCATCCAGACAGAAATAGCGACTTAAATCTCCCATATCA 152
DB 326 IleaIaGlyValIcIyIleaSpGlyArSerLeuValThrlaAsnSerPheArGlyLeu 345
QY 153 ACGACTATTTCAAAAGTTTTCGACCTGAACCAAGAAATACAGCTCTTTGGTGA 212
DB 346 HlStrleuIleaSnleuGlySerIleAProGluProAsnMetThrValleuTrSerGlu 365
QY 213 AATGTGGAAGAAGAGAGCTTATTTATTTATGTCGAACTTCTCTGAAATATGA 272
DB 366 Asnleu---ProGluSerPheIleYsPheCyAlaGluMetSerIleuThrAspSer 384
QY 273 TCCAGGTA-----ACGCTTACCTGTTTATGATGAAAGATTACGGTTA--- 317
DB 385 IlegInTyrgIuAsnAspIleMetArGProIleTyrgIyAspAspTyraIleAla 404
QY 318 -----CGACGACTAGTAAT 332
DB 405 CybCybValSerAlaMetArGValGlyLyAspMetGlnPhePheGlyAlaArGlyCysAsn 424
QY 333 ACCGCTGATGCTTCCCATCATTAATCTACAGATTGTTGATGAGAGGAAAAAATG 392
DB 445 LeuAlaIleCybLeuLeuLeuAlaIleAsnGlyValAspGluIle----- 440
QY 393 TTAATAAGATCATGTGACAGAGGTACACCTAAT---CAACAAATACTTTAAAGCGCTG 449
DB 441 -----LyGlyIleYsValIleProAsnIleGluProIleThrAspGluValleu 457
QY 450 AATTACTAGCGGGGAAAAAATATCTCTGAAATATATACGAT---CAGGTATG 506
DB 458 AspIleYr-----GluIlySvalIlySgluAsnTyPheIlyValleuIlyrMet 473
QY 507 GTTGCTACTATGTAAC 524
DB 474 AlaGlyLeuTyValAsn 479

RESULT 13
US-10-424-599-190153
Sequence 190153, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 190153
LENGTH: 157
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRTJ847_142728C.1.dep
US-10-424-599-190153

Alignment Scores:
Pred. No.: 1.45 Length: 157
Score: 83.50 Matches: 34
Percent Similarity: 35.3% Conservative: 19
Best Local Similarity: 22.7% Mismatches: 46
Query Match: 8.9% Indels: 51
DB: 4 Gaps: 5

US-10-768-093-4 (1-527) x US-10-424-599-190153 (1-157)

QY 81 TTAATTCATATCTTCCTTCATTCGGCAGCCATCCAGACAGAAATAGCGACTTAAATCA 140
DB 4 ValValThrIleCybThrSerGluSerAsnArGlyAlaYsAspGluSerProThrAsnGly 23
QY 141 TTCCAGATATACAGACTATTTCAAAAGTTT----- 173
DB 24 PheIleuValleuSerIleGlyAlaAspPheSerThrGlyLyStryGluIleSerGlu 43
QY 174 -----TTGACCTGAAACACGAATPACAGCTTCTTTGGGAAATGT--- 218
DB 44 LeuValIlySleuIleAlaGluAsnIlyMetCyAlaGluAsnHlSvalMetGluIlySleuVal 63
QY 219 -----GGAAGGAAGAGCTTATTTATTTATGTCGAACTTACGTTCTGAA 266
DB 64 GluAlaAspGluGlyLySgluAspPheIleValTyrgIyValAsnleuThrPhe---Val 82
QY 267 AATGTATCCAGGTACCGCTTAC----- 290
DB 83 AsnleuGluGluValAsnIleTyraSpGlyMetAsnleuAsnGlyGlnIlySproIleMet 102
QY 291 -----CCT 293
DB 103 AlaAsnCybThrPheArGlyValSerAspIlySglValIleuValleuProAlaPro 122
QY 294 GTTATGATGAAGATTATGGTTAGACGACTAGTAATACCGGTGATGCCATCA 353
DB 123 GluAspAspGluAspGlyGlyAsnGlyArGlyIleVal---ThrValSerLeuProArGlu 141
QY 354 ATTAATCTCCAGATTGTGATGAGAAAGG 383
DB 142 GluIleuTyrgIleuIlyAspIlySleuIly 151

RESULT 14
US-10-424-599-250712
Sequence 250712, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 250712
LENGTH: 249
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRTJ847_68422C.1.dep
US-10-424-599-250712

Alignment Scores:
Pred. No.: 1.87 Length: 249
Score: 83.00 Matches: 38
Percent Similarity: 38.4% Conservative: 23
Best Local Similarity: 23.9% Mismatches: 56
Query Match: 8.8% Indels: 42
DB: 4 Gaps: 8

US-10-768-093-4 (1-527) x US-10-424-599-250712 (1-249)

QY 45 CAATTTATATGCTGTATATGAGAAACATTTGTTAATTTCTTCTTCATTC 104
DB 16 LysTyIleuTyrcyTyIYr---HlArGThrIleuHlSlyAsnAlaProSerThrIlePhe 34
QY 105 GGCAGCATGCCAGACAGAAATAGCG-----ACTAAAACTTCCAGTATCAAGACT 158
DB: 4 Gaps: 8

Db 35 AsnAsnHisGlyThrAspGlyAsnAlaTrpProThrSerLeuLeuProLeuAlaLeuAla 54
Oy 159 ATTTCAAAAGTTTTCACCTGACCGAACCAATACACCTTCCTTTT----- 206
Db 55 ValSerIleGlySerLeuAlaLeuGlnProHisPheAsnProSerPheCysAspThrAsp 74
Oy 207 -----GCTGAAAATGTTGGAAAAGAGACGCT----- 233
Db 75 AspArgGlyValGlyValGlyValGlyValGlySerThrGlnValValGlySerGln 94
Oy 234 -----TTATTTATTTAGTGAACCTTAACGTTCCGAAAAGTATCC 275
Db 95 LysGluPheProAspGluLeuLeuGlnAspLeuValValCysGlnAspAsnIleSer 114
Oy 276 CAGTAAACGGTCTACCTGTTTATGATGAGATTATGGGTTAGCA----- 320
Db 115 Leu-----AspTyrAspGluArgTyrIleHisGlyLysProGlnAsnSer 129
Oy 321 -----CGACTAGTAATACCGCTGATGCTTCCCAATCAATATCTACCAAGTTTGAT 374
Db 130 PheHisLysAlaValAsnIleProAspVal-----IleValTyrProArgSerGlu 146
Oy 375 GAGAAAGGAAAAAATGTTAAA-----GATCATGTGCGAGAGTTACACCT 422
Db 147 GluGluValSerLysIleValLysLeuCysAsnSerHisLysValProIleValPro 165

RESULT 15

US-10-501-282-3988 / Sequence 3988, Application US/10501282
Publication No. US20050203280A1
GENERAL INFORMATION:
APPLICANT: MCMICHAEL, JOHN CALHOUN
APPLICANT: ZAGURSKY, ROBERT JOHN
APPLICANT: RUSSELL, DAVID PARRISH
APPLICANT: FLETCHER, LEAH DIANE
TITLE OF INVENTION: ALLOTOCOCUS OBITIDIS OPEN READING FRAMES (ORFS) ENCODING
TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
FILE REFERENCE: AM100780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
CURRENT FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18
PRIOR APPLICATION NUMBER: PCT/US02/36123
PRIOR FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 6653
SOFTWARE: PatentIn version 3.2
SEQ ID NO: 3988
LENGTH: 750
TYPE: PRT
ORGANISM: *Allolococcus obitidis*
US-10-501-282-3988

Alignment Scores:

Pred. No.:	2.51	Length:	750
Score:	83.00	Matches:	38
Percent Similarity:	41.9%	Conservative:	32
Best Local Similarity:	22.8%	Mismatches:	61
Query Match:	8.8%	Indels:	36
DB:	5	Gaps:	8

US-10-768-093-4 (1-527) x US-10-501-282-3988 (1-750)

Oy 93 CTTCCTTCATTCGGCAGCATGCAAGAAATAGCAATAAACTCCAGTATCA 152
Db 335 IleAlaGlyIleGlyLysAspGlyArgHisMetValThrLysAsnSerTyrArgPheLeu 354
Oy 153 ACGACTATTTCAAAAGTTTTCACCTGACCGAACCAATACAGCCTTCCTTTGGTGA 212
Db 355 HisThrLeuSerAsnLeuGlyProAlaProGluProAsnLeuThrValLeuTrpSerGlu 374
Oy 213 AATGTTGA-----AAGGAAGACCTTATATTATTAGTGAACCTTAATGTTCT 263

Db 375 LysLeuProGlnAsnPheLysGluPheCysAlaLysValSerIleAlaSerSerAlaVal 394
Oy 264 GAAATATATCCAGGTAAAGGTTCTACCTGTTTATGATGAGATTATGGGTTAGCA 323
Db 395 GlnTyrGluAsnAspGluIleMetArgProTyrPheGlyAspAspTyrGlyIleAlaCys 414
Oy 324 CTAGTAATATAC----- 344
Db 415 CysValSerIleMetGluIleGlyLysGlnMetGlnPhePheGlyAlaArgAlaAsnLeu 434
Oy 345 TCCCAATCAATATCTACAGATT-----GTGATGAGAAAGGAAAAAATGTGA 395
Db 435 AlaLysAlaLeuLeuTyrAlaIleAsnGlyGlyValAspGlu-----ValThr 450
Oy 396 AAGATCATGTGTCAGAGGTTTACACCTAAT--CAACAAATTAATTAAAGCGCTGAAT 452
Db 451 LysAsp-----GlnValGlyProLysTyrGluProIleThrSerGluTyrLeuAsp 467
Oy 453 TATPCTAGCGGGGAAAAAATATCTCTCGAATATATATACGATCAGTT----- 503
Db 468 TyrAspGluValValGluLysTyr-----AspGlnMetThrGluTrp 481
Oy 504 ATGTTGTTACTATGTAAC 524
Db 482 LeuAlaGlyLeuTyrIleAsn 488

Search completed: April 10, 2006, 10:12:53
Job time : 129.5 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

Om nucleic - protein search, using frame_plus.n2p model

Run on: April 10, 2006, 10:08:49 ; Search time 3.4 Seconds
(without alignments)
966.951 Million cell updates/sec

Title: US-10-768-093-4
Perfect score: 939
Sequence: 1 TTGCACATTCAGATGTTA.....TTGGTTACTACTGTAACATA 527

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 184161 segs, 3191982 residues
Total number of hits satisfying chosen parameters: 368322

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p model -DEV=xlp
-O=/abse/ABSEWEB_pool/US10768093/runat_10042006_090311_4719/app_query.fasta.1
-DB=Published_Applications_AA_New*
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=absn2p
-USER=US10768093_@CCN_1.1.25_@runat_10042006_090311_4719 -NCPU=6 -ICPU=3
-NO_WAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA_New*
1: /SIDSS/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /SIDSS/ptodata/2/pubpaa/US06_NEW_PUB pep:*
3: /SIDSS/ptodata/2/pubpaa/US07_NEW_PUB pep:*
4: /SIDSS/ptodata/2/pubpaa/PCF_NEW_PUB pep:*
5: /SIDSS/ptodata/2/pubpaa/US05_NEW_PUB pep:*
6: /SIDSS/ptodata/2/pubpaa/US10_NEW_PUB pep:*
7: /SIDSS/ptodata/2/pubpaa/US11_NEW_PUB pep:*
8: /SIDSS/ptodata/2/pubpaa/US60_NEW_PUB pep:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	90.5	9.6	268	US-11-195-739-6	Sequence 6, Appl1
2	80.5	8.6	379	US-11-096-568A-6312	Sequence 6312, Ap
3	80.5	8.6	416	US-11-096-568A-6311	Sequence 6311, Ap
4	80.5	8.6	450	US-11-096-568A-6310	Sequence 6310, Ap
5	78.5	8.4	1198	US-10-451-375-4	Sequence 4, Appl1
6	77.5	8.3	1162	US-10-451-375-3	Sequence 3, Appl1
7	75.5	8.0	356	US-11-096-568A-32647	Sequence 32647, A
8	75.5	8.0	361	US-11-096-568A-32646	Sequence 32646, A
9	75.5	8.0	1680	US-10-517-939-362	Sequence 362, App

10	74.5	7.9	400	US-10-793-626-2774	Sequence 2774, Ap
11	73.5	7.8	273	US-11-051-720-1375	Sequence 1375, Ap
12	73.5	7.8	729	US-11-051-720-1373	Sequence 1373, Ap
13	73.5	7.8	738	US-11-051-720-1374	Sequence 1374, Ap
14	73.5	7.8	1006	US-10-793-626-154	Sequence 154, App
15	73.5	7.8	1081	US-11-051-720-1372	Sequence 1372, Ap
16	73.5	7.8	1767	US-10-995-561-911	Sequence 911, App
17	73.5	7.8	1767	US-10-995-561-914	Sequence 914, App
18	73.5	7.8	1806	US-10-995-561-912	Sequence 912, App
19	73.5	7.8	1806	US-10-995-561-915	Sequence 915, App
20	73.5	7.8	1806	US-11-051-720-1446	Sequence 1446, App
21	73.5	7.8	1806	US-11-051-720-1447	Sequence 1447, App
22	73.5	7.8	1818	US-10-995-561-910	Sequence 910, App
23	73.5	7.8	1818	US-10-995-561-913	Sequence 913, App
24	72.5	7.7	465	US-10-485-517-166	Sequence 166, App
25	72.5	7.7	465	US-10-485-517-164	Sequence 164, App
26	70.5	7.5	357	US-09-995-493-178	Sequence 178, App
27	70.5	7.5	357	US-11-194-246-282	Sequence 282, App
28	70.5	7.5	501	US-11-087-099-3301	Sequence 3301, App
29	70.5	7.5	525	US-10-763-712A-108	Sequence 108, App
30	70.5	7.5	898	US-11-099-691-7	Sequence 7, App1
31	70.5	7.5	898	US-11-124-367A-330	Sequence 330, App
32	70.5	7.5	898	US-11-124-367A-333	Sequence 333, App
33	70.5	7.5	905	US-11-087-099-433	Sequence 433, App
34	70	7.5	321	US-11-096-568A-7764	Sequence 7764, App
35	70	7.5	334	US-11-096-568A-7763	Sequence 7763, App
36	70	7.5	365	US-11-096-568A-7762	Sequence 7762, App
37	69.5	7.4	501	US-11-087-099-5784	Sequence 5784, App
38	69.5	7.4	501	US-11-087-099-8619	Sequence 8619, App
39	69	7.3	209	US-11-096-568A-1094	Sequence 1094, App
40	69	7.3	249	US-11-096-568A-1093	Sequence 1093, App
41	69	7.3	1432	US-10-510-386-218	Sequence 218, App
42	68.5	7.3	278	US-10-793-626-2208	Sequence 2208, App
43	68.5	7.3	328	US-10-506-454-335	Sequence 335, App
44	68.5	7.3	340	US-11-057-012-52	Sequence 52, App1
45	68.5	7.3	484	US-11-087-099-1387	Sequence 1387, App

ALIGNMENTS

RESULT 1
US-11-195-739-6
; Sequence 6, Application US/11195739
; Publication No. US20060019324A1
; GENERAL INFORMATION:
; APPLICANT: ALZARI, PEDRO
; APPLICANT: BOITEL, BIRGITTE
; APPLICANT: VILLARINO, ANDREA
; APPLICANT: FERNANDEZ, PABLO
; APPLICANT: COLE, STEWART
; TITLE OF INVENTION: PKB KINASE AND PSTP PHOSPHATASE AND METHODS OF IDENTIFYING
; FILE REFERENCE: 252853US
; CURRENT APPLICATION NUMBER: US/11/195,739
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US/10/892,170
; PRIOR FILING DATE: 2004-07-16
; PRIOR APPLICATION NUMBER: US 60/487,943
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patencin version 3.2
; SEQ ID NO 6
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-11-195-739-6
Alignment Scores: 0.0625
Pred. No.: 90.50
Score: 42.1%
Percent Similarity: 42.1%
Best Local Similarity: 26.8%
Query Match: 9.6%
Length: 268
Matches: 49
Conservative: 28
Mismatch: 54
Indels: 52

DB: 7 Gaps: 10

US-10-768-093-4 (1-527) x US-11-195-739-6 (1-268)

QY 69 AAACATTGGTTAATCTTAATTT-----CTGCTTCATTCGGACCATGCC 116
 DB 91 LysGluValGlyLeuThrAspIleValGluAspGluThrSerPheAsnProPro 110
 QY 117 AGAACAGAAATAGCAGTAAATAAATTCCCA-----GTA 149
 DB 111 GlyThrValIleThrThrIleProAlaProGlyGluValAsnArgGluGluSerVal 130
 QY 150 TCACAGCATTTTCAAAAAGTTTTCACCTGACCGAACAGAAATACAGCCTTTGGT 209
 DB 131 ThrLeuThrIleSerAlaGlyPhe-----PromerProAsnVal-----ValGly 145
 QY 210 GAAAAATGTTGAAAAGAGAGAGCTTTATTATTAGTGTGAATTAAGTCTTCTGAAAT 269
 DB 146 GlnLysValAspAspAlaArgArgLeuLeuGluSerSerAspLeu----- 160
 QY 270 GTATCCAGGTAAACGGTCTACCTGTTTATGATGATGATGATGATGATGATGATGAT 329
 DB 161 -----GluValThrValValGluGluHisIleAspGluValProGluGlyHisValIle 178
 QY 330 AAT-----ACCGCTGATGCTTCCCAATCATTA----- 356
 DB 179 SerGlnGluProGluLysGluThrValGlyValGlyGlnSerValThrLeuThrVal 198
 QY 357 -----ATCTACCAAGATTGTGAT-----GAGAAAGG 383
 DB 199 SerSerGlyProGluLeuValGluValProAspIleArgGlyTrpLysValAspLysAla 218
 QY 384 AAAAAATGTTAAAGATCATGCTGCAGAGTTACACCTTAATCAACAATACTTTTAA 443
 DB 219 ArgGlyGluLeuGluGluValArgGlyPheGluValThrValHisGlnValIleGlyAsnArg 238
 QY 444 GCGCTGAATTACT---AGCGGGGAAAAAAATATCTCTGGATATATTAACGATCAG 500
 DB 239 ValGlyAspTyrAsnProLysGlyGlu-----AlaPro-LysGlySerThrIleGly 255
 QY 501 GTTATGG 507
 DB 255 uilettrp 257

RESULT 2

US-11-096-568A-6312

/ Sequence 6312, Application US/11096568A

/ Publication No. US20060048240A1

/ GENERAL INFORMATION:

/ APPLICANT: Alexandrov, Nikolai et al.

/ TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

/ FILE REFERENCE: 2750-1592PUS2

/ CURRENT APPLICATION NUMBER: US/11/096,568A

/ CURRENT FILING DATE: 2005-04-01

/ NUMBER OF SEQ ID NOS: 34471

/ SEQ ID NO 6312

/ LENGTH: 379

/ TYPE: PR

/ ORGANISM: Glycine max

/ FEATURE:

/ NAME/KEY: misc feature

/ LOCATION: (1)-(379)

/ OTHER INFORMATION: Ceres Seq. ID no. 14314720

US-11-096-568A-6312

Alignment Scores:

Pred. No.: 0.836 Length: 379

Score: 80.50 Matches: 33

Percent Similarity: 34.0% Conservative: 18

Best Local Similarity: 22.0% Mismatches: 48

Query Match: 8.6% Indels: 51

DB: 7 Gaps: 5

US-10-768-093-4 (1-527) x US-11-096-568A-6312 (1-379)

QY 81 TTAATCTTAATCTTGCCTTCAATTCGACGACCATGCCAGAAAGATAATAGCACTAAAAAC 140
 DB 226 ValValThrIleCysThrSerGluSerAsnArgAlaGluAsnGluPheProThrAsnGly 245
 QY 141 TTCACAGTATCAAGCATTTTCAAAAAGTTT----- 173
 DB 246 PheLeuValLeuSerIleGluValAspPheSerThrGlyLysTyrGluIleSerGlu 265
 QY 174 -----TTGCACTGAACACAGAAATACAGCCTTCTTTGGTAAATGTT----- 218
 DB 266 LeuValLysLeuIleAlaGluAsnLysMetValGluAsnHisValMetGluLysLeuVal 285
 QY 219 -----GGAAGAAGAGAGCTTTATTATTATTAAGTGAAGTAACTGTTCTCGAA 266
 DB 286 GluAlaAspGluGlyLysGluLysPheIleValTyrGlyValAsnLeuThrPhe---Val 304
 QY 267 AATGTATCCAGGTAAACGGTCTAC----- 290
 DB 305 AsnLeuGluGluAlaAsnIleTyrAspGlyMetAsnLeuAsnGlyGlnLysProIleMet 324
 QY 291 -----CCT 293
 DB 325 AlaAsnCysThrPheArgGlyValSerAspLysGlyValValLeuValLeuProAlaPro 344
 QY 294 GTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 353
 DB 345 GluAspAspGluAspGlyGlyAsnGlyArgIleVal---ThrValSerLeuProArgGlu 363
 QY 354 ATAATCTTACCAAGATTGTGATGAGAAAGG 383
 DB 364 GluLeuTyrGlnLeuLysAspLysLeuGly 373

RESULT 3

US-11-096-568A-6311

/ Sequence 6311, Application US/11096568A

/ Publication No. US20060048240A1

/ GENERAL INFORMATION:

/ APPLICANT: Alexandrov, Nikolai et al.

/ TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

/ FILE REFERENCE: 2750-1592PUS2

/ CURRENT APPLICATION NUMBER: US/11/096,568A

/ CURRENT FILING DATE: 2005-04-01

/ NUMBER OF SEQ ID NOS: 34471

/ SEQ ID NO 6311

/ LENGTH: 416

/ TYPE: PR

/ ORGANISM: Glycine max

/ FEATURE:

/ NAME/KEY: misc feature

/ LOCATION: (1)-(416)

/ OTHER INFORMATION: Ceres Seq. ID no. 14314719

US-11-096-568A-6311

Alignment Scores:

Pred. No.: 0.851 Length: 416

Score: 80.50 Matches: 33

Percent Similarity: 34.0% Conservative: 18

Best Local Similarity: 22.0% Mismatches: 48

Query Match: 8.6% Indels: 51

DB: 7 Gaps: 5

US-10-768-093-4 (1-527) x US-11-096-568A-6311 (1-416)

QY 81 TTAATCTTAATCTTGCCTTCAATTCGACGACCATGCCAGAAAGATAATAGCACTAAAAAC 140
 DB 263 ValValThrIleCysThrSerGluSerAsnArgAlaGluAsnGluPheProThrAsnGly 282
 QY 141 TTCACAGTATCAAGCATTTTCAAAAAGTTT----- 173

Db 283 PheLeuValLeuSerIleGluAlaAspPheSerThrGlyLysTyrGluIleSerGlu 302
QY 174 -----TTTGCACCTGAAACACGAAATACAGCTTCTTTGGTGAATAATGTT----- 218
Db 303 LeuValIleLeuIleAlaGluAsnIleValMetGluLysLeuVal 322
QY 219 -----GGAAGAAGAGAGCTTATTTATTTAGTGAACCTTAATCTGTTCCGTGA 266
Db 323 GluAlaAspGluGlySerGluPheIleValTyrGlyValAlaMetLeuThrPhe---Val 341
QY 267 AATGATCCAGGTAACGGTCTAC----- 290
Db 342 AsnLeuGluGluAlaAsnIleTyrAspGlyMetAsnLeuAsnGlyGlnLysProIleMet 361
QY 291 -----CCT 293
Db 362 AlaAsnCysThrPheArgGlyValSerAspLysGlyValIleValLeuValLeuProAlaPro 381
QY 294 GTTTATGATGAAGATTATGGTTAGAGACGATGTAATTAACCGCTGATCTTCCCAATCA 353
Db 382 GluAspAspGluAspGlyGlyAsnGlyArgGlyIleVal---ThrValSerLeuProArgGlu 400
QY 354 ATAATCTACCAATCTTGTATGAGAAAGG 383
Db 401 GluLeuTyrGlnLeuLysAspLysLeuGly 410
RESULT 4
US-11-096-568A-6310
/ Sequence 6310, Application US/11096568A
/ Publication No. US20060048240A1
/ GENERAL INFORMATION:
/ APPLICANT: Alexandrov, Nikolai et al.
/ TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
/ TITLE OF INVENTION: Theby
/ FILE REFERENCE: 2750-1592PUS2
/ CURRENT APPLICATION NUMBER: US/11/096, 568A
/ CURRENT FILING DATE: 2005-04-01
/ NUMBER OF SEQ ID NOS: 34471
/ SEQ ID NO 6310
/ LENGTH: 450
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: misc. feature
/ LOCATION: (1)..(450)
/ OTHER INFORMATION: Ceres Seq. ID no. 14314718
US-11-096-568A-6310
Alignment Scores:
Pred. No.: 0.865 Length: 450
Score: 80.50 Matches: 33
Percent Similarity: 34.0% Conservative: 18
Best Local Similarity: 22.0% Mismatches: 48
Query Match: 8.6% Indels: 51
DB: 7 Gaps: 5
US-10-768-093-4 (1-527) x US-11-096-568A-6310 (1-450)
QY 81 TTAATTCTAATCTTGGCTTCTTCCGACCCATGCCAGAACAGAAATAGCCGCTAAAC 140
Db 297 ValValIleThrIleCysThrSerGluSerAsnArgAlaGluAsnGlnPheProThrAsnGly 316
QY 141 TTCACGATCAACGATTTTCAAAAAGTTT----- 173
Db 317 PheLeuValLeuSerIleGluAlaAspPheSerThrGlyLysTyrGluIleSerGlu 336
QY 174 -----TTTGCACCTGAAACACGAAATACAGCTTCTTTGGTGAATAATGTT--- 218
Db 337 LeuValIleLeuIleAlaGluAsnIleValMetValGluAsnIleValMetGluLysLeuVal 356
QY 219 -----GGAAGAAGAGAGCTTATTTATTTAGTGAACCTTAATCTGTTCCGTGA 266
Db 357 GluAlaAspGluGlySerGluAspPheIleValTyrGlyValAlaMetLeuThrPhe---Val 375

QY 267 AATGATCCAGGTAACGGTCTAC----- 290
Db 376 AsnLeuGluGluAlaAsnIleTyrAspGlyMetAsnLeuAsnGlyGlnLysProIleMet 395
QY 291 -----CCT 293
Db 396 AlaAsnCysThrPheArgGlyValSerAspLysGlyValIleValLeuValLeuProAlaPro 415
QY 294 GTTTATGATGAAGATTATGGTTAGAGACGATGTAATTAACCGCTGATCTTCCCAATCA 353
Db 416 GluAspAspGluAspGlyGlyAsnGlyArgGlyIleVal---ThrValSerLeuProArgGlu 434
QY 354 ATAATCTACCAATCTTGTATGAGAAAGG 383
Db 435 GluLeuTyrGlnLeuLysAspLysLeuGly 444
RESULT 5
US-10-451-375-4
/ Sequence 4, Application US/10451375
/ Publication No. US20050261482A1
/ GENERAL INFORMATION:
/ APPLICANT: Bayer AG
/ TITLE OF INVENTION: REGULATION OF HUMAN SERINE-THREONINE PROTEIN KINASE
/ FILE REFERENCE: L10264 Foreign Countries
/ CURRENT APPLICATION NUMBER: US/10/451, 375
/ CURRENT FILING DATE: 2003-07-02
/ PRIOR APPLICATION NUMBER: US 60/259, 215
/ PRIOR FILING DATE: 2001-01-03
/ PRIOR APPLICATION NUMBER: US 60/306, 468
/ PRIOR FILING DATE: 2001-07-20
/ PRIOR APPLICATION NUMBER: US 60/308, 098
/ PRIOR FILING DATE: 2001-07-30
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4
/ LENGTH: 1198
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-451-375-4
Alignment Scores:
Pred. No.: 1.75 Length: 1198
Score: 78.50 Matches: 39
Percent Similarity: 42.0% Conservative: 24
Best Local Similarity: 26.0% Mismatches: 64
Query Match: 8.4% Indels: 23
DB: 6 Gaps: 8
US-10-768-093-4 (1-527) x US-10-451-375-4 (1-1198)
QY 93 CTTCGTTATTCGAGCCATGCCAGAACAGAAATAGCCGCTAAACCTTC----- 143
Db 41 MetThrGlyTyrGlySerHisSerIleValTyrSerGlnSerIleValIleProProSer 60
QY 144 ---CCAGTACACAGCATTTTCAAAAAGTTTTCGACCCGAAACGAAATACAG--- 197
Db 61 GlnProAlaSerThrThrValSerThrSerLeuProValProAsnProSerIleuProTyr 80
QY 198 -----CCTTCCTTTGGTGAATAATGTTGAAAGAAAGAGCTTTATTA 239
Db 81 GluGlnThrIleValPheProGlySerThrGlyHisIleValIleThrSerAlaSerSer 100
QY 240 TTTAGTGTGAAC-----TTAAGTGTCTGTAATAATGTA-----TCCAGGTA 281
Db 101 ThrSerValIleThrGlyGlnValLeuGlyGlyProHisAsnLeuMetArgArgSerThrVal 120
QY 282 ACGGTCTACCTGTTTATGATGAAGATTATGGTTAGAGACGATGTAATTAACCGCTGAT 341
Db 121 SerLeuLeuAspThrTyr---GlnLysCysGlyLeuLysArgGlySerGluGluIleGlu 139
QY 342 CTTCCCAATCAATATATCTACCAATCTTGTATGAGAAAGGAAAGAAATGTTAAAGAT 401
DB: 1 Indels: 11 Gaps: 1

Db 140 AenThSerSeVal-----GlnIleIleGluGluHis---ProPronectIleGlnAsn 156
 QY 402 CATGTGTGACAGAGTTACACCTTAATCAACAATAATCTTTAAAGCGTGAATTACTAGC 461
 Db 157 AenAlaSerGlyAlaThrValAlaThrAlaThrThSerThrAlaThrSerLysAsnSer 176
 QY 462 GGGGAAAAAAATATCTCTCGAATATAT 491
 Db 177 GlySerAsn-----SerGluGlyAspTyr 184

RESULT 6
 US-10-451-375-3
 / Sequence 3, Application US/10451375
 / Publication No. US20050261482A1
 / GENERAL INFORMATION:
 / APPLICANT: Bayer AG
 / TITLE OF INVENTION: REGULATION OF HUMAN SERINE-THREONINE PROTEIN KINASE
 / FILE REFERENCE: LIO264 Foreign Countries
 / CURRENT APPLICATION NUMBER: US/10/451,375
 / CURRENT FILING DATE: 2003-07-02
 / PRIOR APPLICATION NUMBER: US 60/259,215
 / PRIOR FILING DATE: 2001-01-03
 / PRIOR APPLICATION NUMBER: US 60/306,468
 / PRIOR FILING DATE: 2001-07-20
 / PRIOR APPLICATION NUMBER: US 60/308,098
 / PRIOR FILING DATE: 2001-07-30
 / NUMBER OF SEQ ID NOS: 14
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 3
 / LENGTH: 1162
 / TYPE: PRT
 / ORGANISM: Mesocricetus auratus
 / US-10-451-375-3

Alignment Scores:
 Pred. No.: 2.23 Length: 1162
 Score: 77.50 Matches: 40
 Percent Similarity: 42.0% Conservative: 23
 Best Local Similarity: 26.7% Mismatches: 64
 Query Match: 8.3% Indels: 23
 Gaps: 8

US-10-768-093-4 (1-527) x US-10-451-375-3 (1-1162)

QY 93 CTGTGCTTATTCGCGACCCATGCCAGAACAGAAATAGCGACTTAATACTTC----- 143
 Db 34 MetThrGlyTyrGlySerHisSerLysValTyrSerGlnSerLysAsnIleProProSer 53
 QY 144 ---CCGATATCAAGCATATTTTCAAAAGTTTTCGACCTGACCA----- 191
 Db 54 GlnProAlaSerThrThrValSerThrSerLeuProIleProAsnProSerLeuProTyr 73
 QY 192 -----ATACAGCCTTCTTTTGTGAAAGTTGGAAGAGAGAGCTTATTA 239
 Db 74 GluGlnThrIleIleLeuProGlySerThrGlyHisIleValIleThrSerAlaSerSer 93
 QY 240 TTACTGTGTGAAC-----TTACTGTTCTGAAATGTA-----TCCAGGTA 281
 Db 94 ThrSerValThrGlyGlnValLeuGlyGlyProHisAsnLeuMetArgatGSerThrVal 113
 QY 282 ACCGCTTACCTCTTATATGATGTAAGATATAGGTTAGAGACTAGTAATACCGCTGAT 341
 Db 114 SerLeuLeuAspThrTyr---GlnLysCybGlyLeuLysArgLysSerGluGluIleGlu 132
 QY 342 GCTTCCCAATCAATAATCTACAGATGTGTGATGAGAAAGGAAAAAATGTTAAAAAGAT 401
 Db 133 AenThSerSerVal-----GlnIleIleGluGluHis---ProPronectIleGlnAsn 149
 QY 402 CATGTGTGACAGAGTTACACCTTAATCAACAATAATCTTTAAAGCGTGAATTACTAGC 461
 Db 150 AenAlaSerGlyAlaThrValAlaThrAlaThrThSerThrAlaThrSerLysAsnSer 169
 QY 462 GGGGAAAAAAATATCTCTCGAATATAT 491

Db 170 GlySerAsn-----SerGluGlyAspTyr 177

RESULT 7
 US-11-096-568A-32647
 / Sequence 32647, Application US/11096568A
 / Publication No. US20060048240A1
 / GENERAL INFORMATION:
 / APPLICANT: Alexandrov, Nikolai et al.
 / TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
 / FILE REFERENCE: 2750-1592PUS2
 / CURRENT APPLICATION NUMBER: US/11/096,568A
 / CURRENT FILING DATE: 2005-04-01
 / NUMBER OF SEQ ID NOS: 34471
 / SEQ ID NO 32647
 / LENGTH: 356
 / TYPE: PRT
 / ORGANISM: Arabidopsis thaliana
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: (1)..(356)
 / OTHER INFORMATION: Ceres Seq. ID no. 13593732

US-11-096-568A-32647

Alignment Scores:
 Pred. No.: 2.91 Length: 356
 Score: 75.50 Matches: 28
 Percent Similarity: 44.0% Conservative: 16
 Best Local Similarity: 28.0% Mismatches: 23
 Query Match: 8.0% Indels: 33
 Gaps: 6

US-10-768-093-4 (1-527) x US-11-096-568A-32647 (1-356)

QY 258 GTTCCGGAATAATGTATCCAGGTA-----ACGGTCACTCTGTTAT 299
 Db 23 IleProGlnAsnArgAlaGlnIleProGlnTyrPheLysThrAsnValLysProTyrSer 42
 QY 300 GATGAAGATTATGGGTAGACGACTAGTAATACCGCTGATGCTTCCCAATCAATATATC 359
 Db 43 GlnArgLysGlyThrLeuLeuAspProAlaLeuGlnAlaIleGlnAlaIleArgGlnIleIle 62
 QY 360 TACCAATTTGTGACGAAAGGAAAAAATGTTAAAGATCATGCTGCACAGGTTACA 419
 Db 63 -----ThValAsnGlnLysGly-----GlyVala----- 70
 QY 420 CCTAATCAACAATAATCTTTAAAGCGCTGAAT-----TATACTACCGG 464
 Db 71 -----AsnPheLysThrLeuAsnGlnValIleLysSerIleProThrGly 85
 QY 465 GAAAA-----AAATATCTCTCGAATATATATACATGATGATGATGTT 509
 Db 86 AsnLysAsnArgValIleIleLysLeuAlaProGlyValLysAsnGlnLysValThrIle 105

RESULT 8
 US-11-096-568A-32646
 / Sequence 32646, Application US/11096568A
 / Publication No. US20060048240A1
 / GENERAL INFORMATION:
 / APPLICANT: Alexandrov, Nikolai et al.
 / TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
 / FILE REFERENCE: 2750-1592PUS2
 / CURRENT APPLICATION NUMBER: US/11/096,568A
 / CURRENT FILING DATE: 2005-04-01
 / NUMBER OF SEQ ID NOS: 34471
 / SEQ ID NO 32646
 / LENGTH: 361
 / TYPE: PRT
 / ORGANISM: Arabidopsis thaliana
 / FEATURE:
 / NAME/KEY: misc_feature

US-10-768-093-4 (1-527) x US-10-793-626-2774 (1-400)

QY 138 AACCTCCAGATATCAAGACTATTTCACAAAGTTTTCACCTGACCAAGCAATACAG 197
 DB 199 SerPheGlyMetThrGluThrCysSerGln--PheLeuThrAlaSerProGlnMetLeu 217
 QY 198 CCTCTCTTTCGTCAAAATGTTGGAAAG----- 224
 DB 218 LysGluArgPheAspThrValGlyLysProSerGlnAsnValGluValLysIleLysAsn 237
 QY 225 -----GAAAGAGCTTTTATTATTAGTGTGAACCTTAAGTCTCCGAAAT 269
 DB 238 ProAsnAlaIleThrGlyHisGlyGluLeuIleLys-----GlyGluAsn 252
 QY 270 GATATCCAGATACGCTTACCTCTTTATGATGAAGAT----- 308
 DB 253 ValMetAsnGlyIleuThrLeuTyProLysIleuLysAspThrPheAspAsnAspGlyTyr 272
 QY 309 TATGGGTAGGACGACGATCAATACCGCTGATGCTCCCAATCAATATCTACGAGATT 368
 DB 273 PheGlnThrGlyAspIleAlaGluIleAspAspGluGlyTyrValIleIleTyr----- 290
 QY 369 GTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGTCAGAGGTTACACCTAAATCA 428
 DB 291 -----AspArgArgLysAspLeuIleIleSerGlyGlyGluAsnIleTyrPro---Tyr 307
 QY 429 CAATATACCTTTTAAAGCGCTGAATTATCTACGCGGGAAGAAAAATATCTCTCGAATA 488
 DB 308 GlnIleGluThrIleAlaLysAspPheGluGlyIleGluAspAlaValCysValGlyIle 327
 QY 489 TATAACGAT-----CAGCTATGCTGCTGTACTACATTA 521
 DB 328 SerAspAspThrTrpGlyGlnValProIleLeuTyrVal 341

RESULT 11

US-11-051-720-1375
 / Sequence 1375, Application US/11051720
 / Publication No. US20060046257A1
 / GENERAL INFORMATION:
 / APPLICANT: Compugen Ltd
 / TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
 / TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
 / FILE REFERENCE: 1847.1002
 / CURRENT APPLICATION NUMBER: US/11/051.720
 / CURRENT FILING DATE: 2005-01-27
 / NUMBER OF SEQ ID NOS: 1780
 / SEQ ID NO 1375
 / LENGTH: 273
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 US-11-051-720-1375

Alignment Scores:

Pred. No.: 4.58 Length: 273
 Score: 73.50 Matches: 38
 Percent Similarity: 40.9% Conservative: 27
 Best Local Similarity: 23.9% Mismatches: 68
 Query Match: 7.8% Indels: 26
 Gaps: 5

US-10-768-093-4 (1-527) x US-11-051-720-1375 (1-273)
 QY 113 TGGCAGAACGAAATAGCGACTAAATCTCCAGATATCAAGACTATTTCACAA----- 167
 DB 61 CyeThrAsnArgLysAsn-SerLysGlySerAspThrAlaTyrArgValSerLysGlnAl 80
 QY 168 -AGTTTTCGACCGTGAACAGATACAGCCT-----TCCTTTCGTAATAATGTGG 220
 DB 80 aglnLeuSerAlaProThrLysGlnLeuPheProGlyGlyThrPheProGluAspPhe 100
 QY 221 A-----AAGAAGAGCTTTTATTATTATTAGTGTGAACCTTAATGT 259
 DB 100 rIleLeuPheThrValLysProLysLysGlyIleGlnSerPheLeuLeuSerIleTyrAs 120

QY 260 TCCTGAAAAATGATCCAGTACGCTC-----TACCCTGTTAT----- 299
 DB 120 ngluHisGlyIleGlnGlnIleGlyValGluValGlyArgSerProValPheLeuPheG 140
 QY 300 -----GATGAGATTATGGGTTTAGACGACGACTAGTAATAATACCG 337
 DB 140 uAspHisThrGlyLysProAlaProGluAspTyrProLeuPheArgThrValAsnIleAl 160
 QY 338 TGAATGCTCCCAATCAATATATCTACAGATGTTGATGAGAAAGGAAAAATGTTAA 397
 DB 160 aAspGlyLysTrpPheArgValAlaIleSerValGluLysLysThrValThrMetIleVa 180
 QY 398 AGATCATGTGCGAGAGTTACACTTAATCAACAAATTAACCTTTAAAGCGGTGAATTATAC 457
 DB 180 LAspCysLysLysLysThrThrLysProLeuAspArgSerGluArgAlaIleValAspTh 200
 QY 458 TAGCGGGGAAAAAATAATATCTCCGTAATATATATATATATATATATATATATATAT 512
 DB 200 kAsnGlyIleThrValPheGlyThrArgIleLeuAspGluGluValPheGluGly 218

RESULT 12

US-11-051-720-1373
 / Sequence 1373, Application US/11051720
 / Publication No. US20060046257A1
 / GENERAL INFORMATION:
 / APPLICANT: Compugen Ltd
 / TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
 / TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
 / FILE REFERENCE: 1847.1002
 / CURRENT APPLICATION NUMBER: US/11/051.720
 / CURRENT FILING DATE: 2005-01-27
 / NUMBER OF SEQ ID NOS: 1780
 / SEQ ID NO 1373
 / LENGTH: 729
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 US-11-051-720-1373

Alignment Scores:

Pred. No.: 5.58 Length: 729
 Score: 73.50 Matches: 38
 Percent Similarity: 40.9% Conservative: 27
 Best Local Similarity: 23.9% Mismatches: 68
 Query Match: 7.8% Indels: 26
 Gaps: 5

US-10-768-093-4 (1-527) x US-11-051-720-1373 (1-729)
 QY 113 TGGCAGAACGAAATAGCGACTAAATCTCCAGATATCAAGACTATTTCACAA----- 167
 DB 61 CyeThrAsnArgLysAsn-SerLysGlySerAspThrAlaTyrArgValSerLysGlnAl 80
 QY 168 -AGTTTTCGACCGTGAACAGATACAGCCT-----TCCTTTCGTAATAATGTGG 220
 DB 80 aglnLeuSerAlaProThrLysGlnLeuPheProGlyGlyThrPheProGluAspPhe 100
 QY 221 A-----AAGAAGAGCTTTTATTATTATTAGTGTGAACCTTAATGT 259
 DB 100 rIleLeuPheThrValLysProLysLysGlyIleGlnSerPheLeuLeuSerIleTyrAs 120
 QY 260 TCCTGAAAAATGATCCAGTACGCTC-----TACCCTGTTAT----- 299
 DB 120 ngluHisGlyIleGlnGlnIleGlyValGluValGlyArgSerProValPheLeuPheG 140
 QY 300 -----GATGAGATTATGGGTTTAGACGACGACTAGTAATAATACCG 337
 DB 140 uAspHisThrGlyLysProAlaProGluAspTyrProLeuPheArgThrValAsnIleAl 160
 QY 338 TGAATGCTCCCAATCAATATATCTACAGATGTTGATGAGAAAGGAAAAATGTTAA 397
 DB 160 aAspGlyLysTrpPheArgValAlaIleSerValGluLysLysThrValThrMetIleVa 180
 QY 398 AGATCATGTGCGAGAGTTACACTTAATCAACAAATTAACCTTTAAAGCGGTGAATTATAC 457

